

**SELF-PROCESSING PLANTS AND PLANT PARTS****Related Applications**

This application is a continuation-in-part of U.S. Patent Application No. 10/228,063, filed August 27, 2002, which claims priority to Application Serial No. 60/315,281, filed August 27, 2001, each of which is herein incorporated by reference in their entirety.

**Field of the Invention**

The present invention generally relates to the field of plant molecular biology, and more specifically, to the creation of plants that express a processing enzyme which provides a desired characteristic to the plant or products thereof.

**Background of the Invention**

Enzymes are used to process a variety of agricultural products such as wood, fruits and vegetables, starches, juices, and the like. Typically, processing enzymes are produced and recovered on an industrial scale from various sources, such as microbial fermentation (*Bacillus*  $\alpha$ -amylase), or isolation from plants (coffee  $\beta$ -galactosidase or papain from plant parts). Enzyme preparations are used in different processing applications by mixing the enzyme and the substrate under the appropriate conditions of moisture, temperature, time, and mechanical mixing such that the enzymatic reaction is achieved in a commercially viable manner. The methods involve separate steps of enzyme production, manufacture of an enzyme preparation, mixing the enzyme and substrate, and subjecting the mixture to the appropriate conditions to facilitate the enzymatic reaction. A method that reduces or eliminates the time, energy, mixing, capital expenses, and/or enzyme production costs, or results in improved or novel products, would be useful and beneficial. One example of where such improvements are needed is in the area of corn milling.

Today corn is milled to obtain cornstarch and other corn-milling co-products such as corn gluten feed, corn gluten meal, and corn oil. The starch obtained from the process is often further processed into other products such as derivatized starches and sugars, or fermented to make a variety of products including alcohols or lactic acid. Processing of cornstarch often involves the use of enzymes, in particular, enzymes that hydrolyze and convert starch into fermentable sugars

or fructose ( $\alpha$ - and gluco-amylase,  $\alpha$ -glucosidase, glucose isomerase, and the like). The process used commercially today is capital intensive as construction of very large mills is required to process corn on scales required for reasonable cost-effectiveness. In addition the process requires the separate manufacture of starch-hydrolyzing or modifying enzymes and then the machinery to mix the enzyme and substrate to produce the hydrolyzed starch products.

The process of starch recovery from corn grain is well known and involves a wet-milling process. Corn wet-milling includes the steps of steeping the corn kernel, grinding the corn kernel and separating the components of the kernel. The kernels are steeped in a steep tank with a countercurrent flow of water at about 120° F and the kernels remain in the steep tank for 24 to 48 hours. This steepwater typically contains sulfur dioxide at a concentration of about 0.2% by weight. Sulfur dioxide is employed in the process to help reduce microbial growth and also to reduce disulfide bonds in endosperm proteins to facilitate more efficient starch-protein separation. Normally, about 0.59 gallons of steepwater is used per bushel of corn. The steepwater is considered waste and often contains undesirable levels of residual sulfur dioxide.

The steeped kernels are then dewatered and subjected to sets of attrition type mills. The first set of attrition type mills rupture the kernels releasing the germ from the rest of the kernel. A commercial attrition type mill suitable for the wet milling business is sold under the brand name Bauer. Centrifugation is used to separate the germ from the rest of the kernel. A typical commercial centrifugation separator is the Merco centrifugal separator. Attrition mills and centrifugal separators are large expensive items that use energy to operate.

In the next step of the process, the remaining kernel components including the starch, hull, fiber, and gluten are subjected to another set of attrition mills and passed through a set of wash screens to separate the fiber components from the starch and gluten (endosperm protein). The starch and gluten pass through the screens while the fiber does not. Centrifugation or a third grind followed by centrifugation is used to separate the starch from the endosperm protein. Centrifugation produces a starch slurry which is dewatered, then washed with fresh water and dried to about 12% moisture. The substantially pure starch is typically further processed by the use of enzymes.

The separation of starch from the other components of the grain is performed because removing the seed coat, embryo and endosperm proteins allows one to efficiently contact the starch with processing enzymes, and the resulting hydrolysis products are relatively free from contaminants from the other kernel components. Separation also ensures that other components of the grain are effectively recovered and can be subsequently sold as co-products to increase the revenues from the mill.

After the starch is recovered from the wet-milling process it typically undergoes the processing steps of gelatinization, liquefaction and dextrinization for maltodextrin production, and subsequent steps of saccharification, isomerization and refining for the production of glucose, maltose and fructose.

Gelatinization is employed in the hydrolysis of starch because currently available enzymes cannot rapidly hydrolyze crystalline starch. To make the starch available to the hydrolytic enzymes, the starch is typically made into a slurry with water (20-40% dry solids) and heated at the appropriate gelling temperature. For cornstarch this temperature is between 105-110° C. The gelatinized starch is typically very viscous and is therefore thinned in the next step called liquefaction. Liquefaction breaks some of the bonds between the glucose molecules of the starch and is accomplished enzymatically or through the use of acid. Heat-stable endo  $\alpha$ -amylase enzymes are used in this step, and in the subsequent step of dextrinization. The extent of hydrolysis is controlled in the dextrinization step to yield hydrolysis products of the desired percentage of dextrose.

Further hydrolysis of the dextrin products from the liquefaction step is carried out by a number of different exo-amylases and debranching enzymes, depending on the products that are desired. And finally if fructose is desired then immobilized glucose isomerase enzyme is typically employed to convert glucose into fructose.

Dry-mill processes of making fermentable sugars (and then ethanol, for example) from cornstarch facilitate efficient contacting of exogenous enzymes with starch. These processes are less capital intensive than wet-milling but significant cost advantages are still desirable, as often the co-products derived from these processes are not as valuable as those derived from wet-milling. For example, in dry milling corn, the kernel is ground into a powder to facilitate

efficient contact of starch by degrading enzymes. After enzyme hydrolysis of the corn flour the residual solids have some feed value as they contain proteins and some other components. Eckhoff recently described the potential for improvements and the relevant issues related to dry milling in a paper entitled "Fermentation and costs of fuel ethanol from corn with quick-germ process" (Appl. Biochem. Biotechnol., 94: 41 (2001)). The "quick germ" method allows for the separation of the oil-rich germ from the starch using a reduced steeping time.

One example where the regulation and/or level of endogenous processing enzymes in a plant can result in a desirable product is sweet corn. Typical sweet corn varieties are distinguished from field corn varieties by the fact that sweet corn is not capable of normal levels of starch biosynthesis. Genetic mutations in the genes encoding enzymes involved in starch biosynthesis are typically employed in sweet corn varieties to limit starch biosynthesis. Such mutations are in the genes encoding starch synthases and ADP-glucose pyrophosphorylases (such as the sugary and super-sweet mutations). Fructose, glucose and sucrose, which are the simple sugars necessary for producing the palatable sweetness that consumers of edible fresh corn desire, accumulate in the developing endosperm of such mutants. However, if the level of starch accumulation is too high, such as when the corn is left to mature for too long (late harvest) or the corn is stored for an excessive period before it is consumed, the product loses sweetness and takes on a starchy taste and mouthfeel. The harvest window for sweet corn is therefore quite narrow, and shelf-life is limited.

Another significant drawback to the farmer who plants sweet corn varieties is that the usefulness of these varieties is limited exclusively to edible food. If a farmer wanted to forego harvesting his sweet corn for use as edible food during seed development, the crop would be essentially a loss. The grain yield and quality of sweet corn is poor for two fundamental reasons. The first reason is that mutations in the starch biosynthesis pathway cripple the starch biosynthetic machinery and the grains do not fill out completely, causing the yield and quality to be compromised. Secondly, due to the high levels of sugars present in the grain and the inability to sequester these sugars as starch, the overall sink strength of the seed is reduced, which exacerbates the reduction of nutrient storage in the grain. The endosperms of sweet corn variety seeds are shrunken and collapsed, do not undergo proper desiccation, and are susceptible to diseases. The poor quality of the sweet corn grain has further agronomic implications; as poor



seed viability, poor germination, seedling disease susceptibility, and poor early seedling vigor result from the combination of factors caused by inadequate starch accumulation. Thus, the poor quality issues of sweet corn impact the consumer, farmer/grower, distributor, and seed producer.

Thus, for dry-milling, there is a need for a method which improves the efficiency of the process and/or increases the value of the co-products. For wet-milling, there is a need for a method of processing starch that does not require the equipment necessary for prolonged steeping, grinding, milling, and/or separating the components of the kernel. For example, there is a need to modify or eliminate the steeping step in wet milling as this would reduce the amount of waste water requiring disposal, thereby saving energy and time, and increasing mill capacity (kernels would spend less time in steep tanks). There is also a need to eliminate or improve the process of separating the starch-containing endosperm from the embryo.

### Summary of the Invention

The present invention is directed to self-processing plants and plant parts and methods of using the same. The self-processing plant and plant parts of the present invention are capable of expressing and activating enzyme(s) (mesophilic, thermophilic, and/or hyperthermophilic). Upon activation of the enzyme(s) (mesophilic, thermophilic, or hyperthermophilic) the plant or plant part is capable of self-processing the substrate upon which it acts to obtain the desired result.

The present invention is directed to an isolated polynucleotide a) comprising SEQ ID NO: 2, 4, 6, 9, 19, 21, 25, 37, 39, 41, 43, 46, 48, 50, 52, or 59 or the complement thereof, or a polynucleotide which hybridizes to the complement of any one of SEQ ID NO: 2, 4, 6, 9, 19, 21, 25, 37, 39, 41, 43, 46, 48, 50, 52, or 59 under low stringency hybridization conditions and encodes a polypeptide having  $\alpha$ -amylase, pullulanase,  $\alpha$ -glucosidase, glucose isomerase, or glucoamylase activity or b) encoding a polypeptide comprising SEQ ID NO: 10, 13, 14, 15, 16, 18, 20, 24, 26, 27, 28, 29, 30, 33, 34, 35, 36, 38, 40, 42, 44, 45, 47, 49, or 51 or an enzymatically active fragment thereof. Preferably, the isolated polynucleotide encodes a fusion polypeptide comprising a first polypeptide and a second peptide, wherein said first polypeptide has  $\alpha$ -amylase, pullulanase,  $\alpha$ -glucosidase, glucose isomerase, or glucoamylase activity. Most preferably, the second peptide comprises a signal sequence peptide, which may target the first

polypeptide to a vacuole, endoplasmic reticulum, chloroplast, starch granule, seed or cell wall of a plant. For example, the signal sequence may be an N-terminal signal sequence from *waxy*, an N-terminal signal sequence from  $\gamma$ -zein, a starch binding domain, or a C-terminal starch binding domain. Polynucleotides that hybridize to the complement of any one of SEQ ID NO: 2, 9, or 52 under low stringency hybridization conditions and encodes a polypeptide having  $\alpha$ -amylase activity; to the complement of SEQ ID NO: 4 or 25 under low stringency hybridization conditions and encodes a polypeptide having pullulanase activity; to the complement of SEQ ID NO: 6 and encodes a polypeptide having  $\alpha$ -glucosidase activity; to the complement of any one of SEQ ID NO: 19, 21, 37, 39, 41, or 43 under low stringency hybridization conditions and encodes a polypeptide having glucose isomerase activity; to the complement of any one of SEQ ID NO: 46, 48, 50, or 59 under low stringency hybridization conditions and encodes a polypeptide having glucoamylase activity are further encompassed.

The present invention is also directed to an isolated polynucleotide a) comprising SEQ ID NO: 61, 63, 65, 79, 81, 83, 85, 87, 89, 91, 93, 94, 95, 96, 97, 99, 108, and 110 or the complement thereof, or a polynucleotide which hybridizes to the complement of any one of SEQ ID NO: 61, 63, 65, 79, 81, 83, 85, 87, 89, 91, 93, 94, 95, 96, 97, 99, 108, or 110 under low stringency hybridization conditions and encodes a polypeptide having xylanase, cellulase, glucanase, beta glucosidase, esterase or phytase activity b) encoding a polypeptide comprising SEQ ID NO: 62, 64, 66, 70, 80, 82, 84, 86, 88, 90, 92, 109, or 111 or an enzymatically active fragment thereof. The isolated polynucleotide may encode a fusion polypeptide comprising a first polypeptide and a second peptide, wherein said first polypeptide has xylanase, cellulase, glucanase, beta glucosidase, protease, or phytase activity. The second peptide may comprises a signal sequence peptide, which may target the first polypeptide to a vacuole, endoplasmic reticulum, chloroplast, starch granule, seed or cell wall of a plant. For example, the signal sequence may be an N-terminal signal sequence from *waxy*, an N-terminal signal sequence from  $\gamma$ -zein, a starch binding domain, or a C-terminal starch binding domain.

Exemplary xylanases provided and useful in the invention include those encoded by SEQ ID NO: 61, 63, or 65. An exemplary protease, namely bromelain, encoded by SEQ ID NO: 69 is also provided. Exemplary cellulases include cellobiohydrolase I and II as provided herein and

encoded by SEQ ID NO: 79, 81, 93, and 94. An exemplary glucanase is provided as 6GP1 described herein encoded by SEQ ID NO: 85. Exemplary beta glucosidases include beta glucosidase 2 and D, as described herein and encoded by SEQ ID NO: 96 and 97. An exemplary esterase is also provided, namely ferulic acid esterase as encoded by SEQ ID NO: 99. And, an exemplary phytase, Nov9X as encoded by SEQ ID NO: 109-112 is also provided.

Also included are expression cassettes comprising a polynucleotide a) having SEQ ID NO: 2, 4, 6, 9, 19, 21, 25, 37, 39, 41, 43, 46, 48, 50, 52, or 59 or the complement thereof, or a polynucleotide which hybridizes to the complement of any one of SEQ ID NO: 2, 4, 6, 9, 19, 21, 25, 37, 39, 41, 43, 46, 48, 50, 52, or 59 or under low stringency hybridization conditions and encodes an polypeptide having  $\alpha$ -amylase, pullulanase,  $\alpha$ -glucosidase, glucose isomerase, or glucoamylase activity or b) encoding a polypeptide comprising SEQ ID NO: 10, 13, 14, 15, 16, 18, 20, 24, 26, 27, 28, 29, 30, 33, 34, 35, 36, 38, 40, 42, 44, 45, 47, 49, or 51, or an enzymatically active fragment thereof. The expression cassette further comprises a promoter operably linked to the polynucleotide, such as an inducible promoter, tissue-specific promoter, or preferably an endosperm-specific promoter. Preferably, the endosperm-specific promoter is a maize  $\gamma$ -zein promoter or a maize ADP-gpp promoter or a maize Q promoter or a rice glutelin-1 promoter. In a preferred embodiment, the promoter comprises SEQ ID NO: 11 or SEQ ID NO: 12 or SEQ ID NO: 67 or SEQ ID NO: 98. Moreover, in another preferred embodiment the polynucleotide is oriented in sense orientation relative to the promoter. The expression cassette of the present invention may further encode a signal sequence which is operably linked to the polypeptide encoded by the polynucleotide. The signal sequence preferably targets the operably linked polypeptide to a vacuole, endoplasmic reticulum, chloroplast, starch granule, seed or cell wall of a plant. The signal sequences include an N-terminal signal sequence from *waxy*, an N-terminal signal sequence from  $\gamma$ -zein, or a starch binding domain.

Moreover, an expression cassette comprising a polynucleotide a) having SEQ ID NO: 61, 63, 65, 79, 81, 83, 85, 87, 89, 91, 93, 94, 95, 96, 97, 99, 108, and 110 or the complement thereof, or a polynucleotide which hybridizes to the complement of any one of SEQ ID NO: 61, 63, 65, 79, 81, 83, 85, 87, 89, 91, 93, 94, 95, 96, 97, 99, 108, and 110 or under low stringency hybridization conditions and encodes an polypeptide having xylanase, cellulase, glucanase, beta

glucosidase, esterase or phytase activity or b) encoding a polypeptide comprising SEQ ID NO: 62, 64, 66, 70, 80, 82, 84, 86, 88, 90, 92, 109, or 111, or an enzymatically active fragment thereof. The expression cassette further comprises a promoter operably linked to the polynucleotide, such as an inducible promoter, tissue-specific promoter, or preferably an endosperm-specific promoter. The endosperm-specific promoter may be a maize  $\gamma$ -zein promoter or a maize ADP-gpp promoter or a maize Q promoter promoter or a rice glutelin-1 promoter. In an embodiment, the promoter comprises SEQ ID NO: 11 or SEQ ID NO: 12 or SEQ ID NO: 67 or SEQ ID NO: 98. Moreover, in another embodiment the polynucleotide is oriented in sense orientation relative to the promoter. The expression cassette of the present invention may further encode a signal sequence which is operably linked to the polypeptide encoded by the polynucleotide. The signal sequence preferably targets the operably linked polypeptide to a vacuole, endoplasmic reticulum, chloroplast, starch granule, seed or cell wall of a plant. The signal sequences include an N-terminal signal sequence from *waxy*, an N-terminal signal sequence from  $\gamma$ -zein, or a starch binding domain.

The present invention is further directed to a vector or cell comprising the expression cassettes of the present invention. The cell may be selected from the group consisting of an *Agrobacterium*, a monocot cell, a dicot cell, a Liliopsida cell, a Panicoideae cell, a maize cell, and a cereal cell, such as a rice cell.

Moreover, the present invention encompasses a plant stably transformed with the vectors of the present invention. A plant stably transformed with a vector comprising an  $\alpha$ -amylase having an amino acid sequence of any of SEQ ID NO: 1, 10, 13, 14, 15, 16, 33, 35 or 88 or encoded by a polynucleotide comprising any of SEQ ID NO: 2, 9, or 87 is provided.

In another embodiment, a plant stably transformed with a vector comprising a pullulanase having an amino acid sequence of any of SEQ ID NO: 24 or 34, or encoded by a polynucleotide comprising any of SEQ ID NO: 4 or 25 is provided. A plant stably transformed with a vector comprising an  $\alpha$ -glucosidase having an amino acid sequence of any of SEQ ID NO: 26 or 27, or encoded by a polynucleotide comprising SEQ ID NO: 6 is further provided. A plant stably transformed with a vector comprising an glucose isomerase having an amino acid sequence of any of SEQ ID NO: 18, 20, 28, 29, 30, 38, 40, 42, or 44, or encoded by a polynucleotide

comprising any of SEQ ID NO:19, 21, 37, 39, 41, or 43 is further described herein. In another embodiment, a plant stably transformed with a vector comprising a glucose amylase having an amino acid sequence of any of SEQ ID NO: 45, 47, or 49, or encoded by a polynucleotide comprising any of SEQ ID NO:46, 48, 50, or 59 is described.

An additional embodiment provides a plant stably transformed with a vector comprising a xylanase having an amino acid sequence of any of SEQ ID NO: 62, 64 or 66, or encoded by a polynucleotide comprising any of SEQ ID NO: 61, 63, or 65. A plant stably transformed with a vector comprising a protease is also provided. The protease may be bromelain having an amino acid sequence as set forth in SEQ ID NO: 70, or encoded by a polynucleotide having SEQ ID NO: 69. In another embodiment, a plant stably transformed with a vector comprising a cellulase is provided. The cellulase may be a cellobiohydrolase encoded by a polynucleotide comprising any of SEQ ID NO: 79, 80, 81, 82, 93 or 94.

An additional embodiment provides a plant stably transformed with a vector comprising a glucanase, such as an endoglucanase. The endoglucanase may be endoglucanase I which has an amino acid sequence as in SEQ ID NO: 84, or encoded by a polynucleotide comprising SEQ ID NO: 83. A plant stably transformed with a vector comprising a beta glucosidase is also provided. The beta glucosidase is may be beta glucosidase 2 or beta glucosidase D, which have an amino acid sequence set forth in SEQ ID NO: 90 or 92, or encoded by a polynucleotide having SEQ ID NO: 89 or 91. In another embodiment, a plant stably transformed with a vector comprising an esterase is provided. The esterase may be a ferulic acid esterase encoded by a polynucleotide comprising SEQ ID NO: 99.

Plant products, such as seed, fruit or grain from the stably transformed plants of the present invention are further provided.

In another embodiment, the invention is directed to a transformed plant, the genome of which is augmented with a recombinant polynucleotide encoding at least one processing enzyme operably linked to a promoter sequence, the sequence of which polynucleotide is optimized for expression in the plant. The plant may be a monocot, such as maize or rice, or a dicot. The plant may be a cereal plant or a commercially grown plant. The processing enzyme is selected from the group consisting of an  $\alpha$ -amylase, glucoamylase, glucose isomerase, glucanase,  $\beta$ -

amylase,  $\alpha$ -glucosidase, isoamylase, pullulanase, neo-pullulanase, iso-pullulanase, amylopullulanase, cellulase, exo-1,4- $\beta$ -cellobiohydrolase, exo-1,3- $\beta$ -D-glucanase,  $\beta$ -glucosidase, endoglucanase, L-arabinase,  $\alpha$ -arabinosidase, galactanase, galactosidase, mannanase, mannosidase, xylanase, xylosidase, protease, glucanase, xylanase, , esterase, phytase, and lipase. The processing enzyme is a starch-processing enzyme selected from the group consisting of  $\alpha$ -amylase, glucoamylase, glucose isomerase,  $\beta$ -amylase,  $\alpha$ -glucosidase, isoamylase, pullulanase, neo-pullulanase, iso-pullulanase, and amylopullulanase. The enzyme may be selected from  $\alpha$ -amylase, glucoamylase, glucose isomerase, glucose isomerase,  $\alpha$ -glucosidase, and pullulanase. The processing enzyme may be hyperthermophilic. In accordance with this aspect of the invention, the enzyme may be a non-starch degrading enzyme selected from the group consisting of protease, glucanase, xylanase, esterase, phytase, cellulase, beta glucosidase, and lipase. Such enzymes may be hyperthermophilic. In an embodiment, the enzyme accumulates in the vacuole, endoplasmic reticulum, chloroplast, starch granule, seed or cell wall of a plant. Moreover, in another embodiment, the genome of plant may be further augmented with a second recombinant polynucleotide comprising a non-hyperthermophilic enzyme.

In another aspect of the invention, provided is a transformed plant, the genome of which is augmented with a recombinant polynucleotide encoding at least one processing enzyme selected from the group consisting of  $\alpha$ -amylase, glucoamylase, glucose isomerase,  $\alpha$ -glucosidase, pullulanase, xylanase, cellulase, protease, glucanase, beta glucosidase, esterase, phytase or lipase operably linked to a promoter sequence, the sequence of which polynucleotide is optimized for expression in the plant.

Another embodiment is directed to a transformed maize plant, the genome of which is augmented with a recombinant polynucleotide encoding at least one processing enzyme selected from the group consisting of  $\alpha$ -amylase, glucoamylase, glucose isomerase,  $\alpha$ -glucosidase, pullulanase, xylanase, cellulase, protease, glucanase, phytase, beta glucosidase, esterase, or lipase operably linked to a promoter sequence, the sequence of which polynucleotide is optimized for expression in the maize plant.

A transformed plant, the genome of which is augmented with a recombinant polynucleotide having SEQ ID NO: 83 operably linked to a promoter and to a signal sequence is

provided. Additionally, a transformed plant, the genome of which is augmented with a recombinant polynucleotide having the SEQ ID NO: 93 or 94 operably linked to a promoter and to a signal sequence is described. In another embodiment, a transformed plant, the genome of which is augmented with a recombinant polynucleotide having SEQ ID NO: 95, operably linked to a promoter and to a signal sequence. Moreover, a transformed plant, the genome of which is augmented with a recombinant polynucleotide having SEQ ID NO: 96 is described. Also described is a transformed plant, the genome of which is augmented with a recombinant polynucleotide having SEQ ID NO: 97. Also described is a transformed plant, the genome of which is augmented with a recombinant polypeptide having SEQ ID NO: 99.

Products of the transformed plants are further envisioned herein. The product for example, include seed, fruit, or grain. The product may alternatively be the processing enzyme, starch or sugar.

A plant obtained from a stably transformed plant of the present invention is further described. In this aspect, the plant may be a hybrid plant or an inbred plant.

A starch composition is a further embodiment of the invention comprising at least one processing enzyme which is a protease, glucanase, or esterase.

Grain is another embodiment of the invention comprising at least one processing enzyme, which is an  $\alpha$ -amylase, pullulanase,  $\alpha$ -glucosidase, glucoamylase, glucose isomerase, xylanase, cellulase, glucanase, beta glucosidase, esterase, protease, lipase or phytase.

In another embodiment, a method of preparing starch granules, comprising treating grain which comprises at least one non-starch processing enzyme under conditions which activate the at least one enzyme, yielding a mixture comprising starch granules and non-starch degradation products, wherein the grain is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one enzyme; and separating starch granules from the mixture is provided. Therein, the enzyme may be a protease, glucanase, xylanase, phytase, lipase, beta glucosidase, cellulase or esterase. Moreover, the enzyme is preferably hyperthermophilic. The grain may be cracked grain and/or may be treated under low or high moisture conditions. Alternatively, the grain may be treated with sulfur dioxide. The present invention may further comprise separating non-starch products from the mixture. The starch products and non-starch products obtained by this method are further described.

In yet another embodiment, a method to produce hypersweet corn comprising treating transformed corn or a part thereof, the genome of which is augmented with and expresses in the endosperm an expression cassette encoding at least one starch-degrading or starch-isomerizing enzyme, under conditions which activate the at least one enzyme so as to convert polysaccharides in the corn into sugar, yielding hypersweet corn is provided. The expression cassette may further comprises a promoter operably linked to the polynucleotide encoding the enzyme. The promoter may be a constitutive promoter, seed-specific promoter, or endosperm-specific promoter, for example. The enzyme may be hyperthermophilic and may be an  $\alpha$ -amylase. The expression cassette used herein may further comprise a polynucleotide which encodes a signal sequence operably linked to the at least one enzyme. The signal sequence may direct the enzyme to the apoplast or the endoplasmic reticulum, for example. The enzyme comprises any one of SEQ ID NO: 13, 14, 15, 16, 33, or 35. The enzyme may also comprise SEQ ID NO: 87.

In a most preferred embodiment, a method of producing hypersweet corn comprising treating transformed corn or a part thereof, the genome of which is augmented with and expresses in the endosperm an expression cassette encoding an  $\alpha$ -amylase, under conditions which activate the at least one enzyme so as to convert polysaccharides in the corn into sugar, yielding hypersweet corn is described. The enzyme may be hyperthermophilic and the hyperthermophilic  $\alpha$ -amylase may comprise the amino acid sequence of any of SEQ ID NO: 10, 13, 14, 15, 16, 33, or 35, or an enzymatically active fragment thereof having  $\alpha$ -amylase activity. The enzyme comprise SEQ ID NO: 87.

A method to prepare a solution of hydrolyzed starch product comprising; treating a plant part comprising starch granules and at least one processing enzyme under conditions which activate the at least one enzyme thereby processing the starch granules to form an aqueous solution comprising hydrolyzed starch product, wherein the plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one starch processing enzyme; and collecting the aqueous solution comprising the hydrolyzed starch product is described herein. The hydrolyzed starch product may comprise a dextrin, maltooligosaccharide, glucose and/or mixtures thereof. The enzyme may be  $\alpha$ -amylase,  $\alpha$ -glucosidase, glucoamylase, pullulanase,



amylopullulanase, glucose isomerase, or any combination thereof. Moreover, the enzyme may be hyperthermophilic. In another aspect, the genome of the plant part may be further augmented with an expression cassette encoding a non-hyperthermophilic starch processing enzyme. The non-hyperthermophilic starch processing enzyme may be selected from the group consisting of amylase, glucoamylase,  $\alpha$ -glucosidase, pullulanase, glucose isomerase, or a combination thereof. In yet another aspect, the processing enzyme is preferably expressed in the endosperm. The plant part may be grain, and from corn, wheat, barley, rye, oat, sugar cane or rice. The at least one processing enzyme is operably linked to a promoter and to a signal sequence that targets the enzyme to the starch granule or the endoplasmic reticulum, or to the cell wall. The method may further comprise isolating the hydrolyzed starch product and/or fermenting the hydrolyzed starch product.

In another aspect of the invention, a method of preparing hydrolyzed starch product comprising treating a plant part comprising starch granules and at least one starch processing enzyme under conditions which activate the at least one enzyme thereby processing the starch granules to form an aqueous solution comprising a hydrolyzed starch product, wherein the plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding at least one  $\alpha$ -amylase; and collecting the aqueous solution comprising hydrolyzed starch product is described. The  $\alpha$ -amylase may be hyperthermophilic and the hyperthermophilic  $\alpha$ -amylase comprises the amino acid sequence of any of SEQ ID NO: 1, 10, 13, 14, 15, 16, 33, or 35, or an active fragment thereof having  $\alpha$ -amylase activity. The expression cassette may comprise a polynucleotide selected from any of SEQ ID NO: 2, 9, 46, or 52, a complement thereof, or a polynucleotide that hybridizes to any of SEQ ID NO: 2, 9, 46, or 52 under low stringency hybridization conditions and encodes a polypeptide having  $\alpha$ -amylase activity. Moreover, the invention further provides for the genome of the transformed plant further comprising a polynucleotide encoding a non-thermophilic starch-processing enzyme. Alternatively, the plant part may be treated with a non-hyperthermophilic starch-processing enzyme.

The present invention is further directed to a transformed plant part comprising at least one starch-processing enzyme present in the cells of the plant, wherein the plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette

encoding the at least one starch processing enzyme. Preferably, the enzyme is a starch-processing enzyme selected from the group consisting of  $\alpha$ -amylase, glucoamylase, glucose isomerase,  $\beta$ -amylase,  $\alpha$ -glucosidase, isoamylase, pullulanase, neo-pullulanase, iso-pullulanase, and amylopullulanase. Moreover, the enzyme may be hyperthermophilic. The plant may be any plant, such as corn or rice for example.

Another embodiment of the invention is a transformed plant part comprising at least one non-starch processing enzyme present in the cell wall or the cells of the plant, wherein the plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one non-starch processing enzyme or at least one non-starch polysaccharide processing enzyme. The enzyme may be hyperthermophilic. Moreover, the non-starch processing enzyme may be a protease, glucanase, xylanase, esterase, phytase, beta glucosidase, cellulase or lipase. The plant part can be any plant part, but preferably is an ear, seed, fruit, grain, stover, chaff, or bagasse.

The present invention is also directed to transformed plant parts. For example, a transformed plant part comprising an  $\alpha$ -amylase having an amino acid sequence of any of SEQ ID NO: 1, 10, 13, 14, 15, 16, 33, or 35, or encoded by a polynucleotide comprising any of SEQ ID NO: 2, 9, 46, or 52, a transformed plant part comprising an  $\alpha$ -glucosidase having an amino acid sequence of any of SEQ ID NO: 5, 26 or 27, or encoded by a polynucleotide comprising SEQ ID NO:6, a transformed plant part comprising a glucose isomerase having the amino acid sequence of any one of SEQ ID NO: 28, 29, 30, 38, 40, 42, or 44, or encoded by a polynucleotide comprising any one of SEQ ID NO: 19, 21, 37, 39, 41, or 43, a transformed plant part comprising a glucoamylase having the amino acid sequence of SEQ ID NO:45 or SEQ ID NO:47, or SEQ ID NO:49, or encoded by a polynucleotide comprising any of SEQ ID NO: 46, 48, 50, or 59, and a transformed plant part comprising a pullulanase encoded by a polynucleotide comprising any of SEQ ID NO: 4 or 25 are described.

The present invention is also directed to transformed plant parts. For example, a transformed plant part comprising a xylanase having an amino acid sequence of any of SEQ ID NO: 62, 64 or 66, or encoded by a polynucleotide comprising any of SEQ ID NO: 61, 63, or 65.

A transformed plant part comprising a protease is also provided. The protease may be bromelain having an amino acid sequence as set forth in SEQ ID NO: 70, or encoded by a polynucleotide having SEQ ID NO: 69. In another embodiment, a transformed plant part comprising a cellulase is provided. The cellulase may be a cellobiohydrolase encoded by a polynucleotide comprising any of SEQ ID NO: 79, 80, 81, 82, 93 or 94.

An additional embodiment provides a transformed plant part a glucanase, such as an endoglucanase. The endoglucanase may be endoglucanase I which has an amino acid sequence as in SEQ ID NO: 84, or encoded by a polynucleotide comprising SEQ ID NO: 83. A transformed plant part comprising a beta glucosidase is also provided. The beta glucosidase is may be beta glucosidase 2 or beta glucosidase D, which have an amino acid sequence set forth in SEQ ID NO: 90 or 92, or encoded by a polynucleotide having SEQ ID NO: 89 or 91. In another embodiment, a transformed plant part comprising an esterase is provided. The esterase may be a ferulic acid esterase encoded by a polynucleotide comprising SEQ ID NO: 99.

Another embodiment is a method of converting starch in the transformed plant part comprising activating the starch processing enzyme contained therein. The starch, dextrin, maltooligosaccharide or sugar produced according to this method is further described.

The present invention further describes a method of using a transformed plant part comprising at least one non-starch processing enzyme in the cell wall or the cell of the plant part, comprising treating a transformed plant part comprising at least one non-starch polysaccharide processing enzyme under conditions so as to activate the at least one enzyme thereby digesting non-starch polysaccharide to form an aqueous solution comprising oligosaccharide and/or sugars, wherein the plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one non-starch polysaccharide processing enzyme; and collecting the aqueous solution comprising the oligosaccharides and/or sugars. The non-starch polysaccharide processing enzyme may be hyperthermophilic.

A method of using transformed seeds comprising at least one processing enzyme, comprising treating transformed seeds which comprise at least one protease or lipase under conditions so as the activate the at least one enzyme yielding an aqueous mixture comprising amino acids and fatty acids, wherein the seed is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one enzyme; and

collecting the aqueous mixture. The amino acids, fatty acids or both are preferably isolated. The at least one protease or lipase may be hyperthermophilic.

A method to prepare ethanol comprising treating a plant part comprising at least one polysaccharide processing enzyme under conditions to activate the at least one enzyme thereby digesting polysaccharide to form oligosaccharide or fermentable sugar, wherein the plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one polysaccharide processing enzyme; and incubating the fermentable sugar under conditions that promote the conversion of the fermentable sugar or oligosaccharide into ethanol. The

plant part may be a grain, fruit, seed, stalks, wood, vegetable or root. The plant part may be obtained from a plant selected from the group consisting of oats, barley, wheat, berry, grapes, rye, corn, rice, potato, sugar beet, sugar cane, pineapple, grasses and trees.

In another preferred embodiment, the polysaccharide processing enzyme is  $\alpha$ -amylase, glucoamylase,  $\alpha$ -glucosidase, glucose isomerase, pullulanase, or a combination thereof.

A method to prepare ethanol comprising treating a plant part comprising at least one enzyme selected from the group consisting of  $\alpha$ -amylase, glucoamylase,  $\alpha$ -glucosidase, glucose isomerase, or pullulanase, or a combination thereof, with heat for an amount of time and under conditions to activate the at least one enzyme thereby digesting polysaccharide to form fermentable sugar, wherein the plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one enzyme; and incubating the fermentable sugar under conditions that promote the conversion of the fermentable sugar into ethanol is provided. The at least one enzyme may be hyperthermophilic or mesophilic.

In another embodiment, a method to prepare ethanol comprising treating a plant part comprising at least one non-starch processing enzyme under conditions to activate the at least one enzyme thereby digesting non-starch polysaccharide to oligosaccharide and fermentable sugar, wherein the plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one enzyme; and incubating the fermentable sugar under conditions that promote the conversion of the fermentable sugar into

ethanol is provided. The non-starch processing enzyme may be a xylanase, cellulase, glucanase, beta glucosidase, protease, esterase, lipase or phytase.

A method to prepare ethanol comprising treating a plant part comprising at least one enzyme selected from the group consisting of  $\alpha$ -amylase, glucoamylase,  $\alpha$ -glucosidase, glucose isomerase, or pullulanase, or a combination thereof, under conditions to activate the at least one enzyme thereby digesting polysaccharide to form fermentable sugar, wherein the plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one enzyme; and incubating the fermentable sugar under conditions that promote the conversion of the fermentable sugar into ethanol is further provided. The enzyme may be hyperthermophilic.

Moreover, a method to produce a sweetened farinaceous food product without adding additional sweetener comprising treating a plant part comprising at least one starch processing enzyme under conditions which activate the at least one enzyme, thereby processing starch granules in the plant part to sugars so as to form a sweetened product, wherein the plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one enzyme; and processing the sweetened product into a farinaceous food product is described. The farinaceous food product may be formed from the sweetened product and water. Moreover, the farinaceous food product may contain malt, flavorings, vitamins, minerals, coloring agents or any combination thereof. The at least one enzyme may be hyperthermophilic. The enzyme may be selected from  $\alpha$ -amylase,  $\alpha$ -glucosidase, glucoamylase, pullulanase, glucose isomerase, or any combination thereof. The plant may further be selected from the group consisting of soybean, rye, oats, barley, wheat, corn, rice and sugar cane. The farinaceous food product may be a cereal food, a breakfast food, a ready to eat food, or a baked food. The processing may include baking, boiling, heating, steaming, electrical discharge or any combination thereof.

The present invention is further directed to a method to sweeten a starch-containing product without adding sweetener comprising treating starch comprising at least one starch processing enzyme under conditions to activate the at least one enzyme thereby digesting the starch to form a sugar to form sweetened starch, wherein the starch is obtained from a

transformed plant, the genome of which is augmented with an expression cassette encoding the at least one enzyme; and adding the sweetened starch to a product to produce a sweetened starch containing product. The transformed plant may be selected from the group consisting of corn, soybean, rye, oats, barley, wheat, rice and sugar cane. The at least one enzyme may be hyperthermophilic. The at least one enzyme may be  $\alpha$ -amylase,  $\alpha$ -glucosidase, glucoamylase, pullulanase, glucose isomerase, or any combination thereof.

A farinaceous food product and sweetened starch-containing product is provided for herein.

The invention is also directed to a method to sweeten a polysaccharide-containing fruit or vegetable comprising treating a fruit or vegetable comprising at least one polysaccharide processing enzyme under conditions which activate the at least one enzyme, thereby processing the polysaccharide in the fruit or vegetable to form sugar, yielding a sweetened fruit or vegetable, wherein the fruit or vegetable is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one polysaccharide processing enzyme. The fruit or vegetable is selected from the group consisting of potato, tomato, banana, squash, peas, and beans. The at least one enzyme may be hyperthermophilic.

The present invention is further directed to a method of preparing an aqueous solution comprising sugar comprising treating starch granules obtained from the plant part under conditions which activate the at least one enzyme, thereby yielding an aqueous solution comprising sugar.

Another embodiment is directed to a method of preparing starch derived products from grain that does not involve wet or dry milling grain prior to recovery of starch-derived products comprising treating a plant part comprising starch granules and at least one starch processing enzyme under conditions which activate the at least one enzyme thereby processing the starch granules to form an aqueous solution comprising dextrans or sugars, wherein the plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one starch processing enzyme; and collecting the aqueous solution comprising the starch derived product. The at least one starch processing enzyme may be hyperthermophilic.

A method of isolating an  $\alpha$ -amylase, glucoamylase, glucose isomerase,  $\alpha$ -glucosidase, and pullulanase comprising culturing a transformed plant containing the  $\alpha$ -amylase, glucoamylase, glucose isomerase,  $\alpha$ -glucosidase, or pullulanase and isolating the  $\alpha$ -amylase, glucoamylase, glucose isomerase,  $\alpha$ -glucosidase or pullulanase therefrom is further provided. Also provided is a method of isolating a xylanase, cellulase, glucanase, beta glucosidase, protease, esterase, phytase or lipase comprising culturing a transformed plant containing the xylanase, cellulase, glucanase, beta glucosidase, protease, esterase, phytase or lipase and isolating the xylanase, cellulase, glucanase, esterase, beta glucosidase, protease, esterase, phytase or lipase .

A method of preparing maltodextrin comprising mixing transgenic grain with water, heating said mixture, separating solid from the dextrin syrup generated, and collecting the maltodextrin. The transgenic grain comprises at least one starch processing enzyme. The starch processing enzyme may be  $\alpha$ -amylase, glucoamylase,  $\alpha$ -glucosidase, and glucose isomerase. Moreover, maltodextrin produced by the method is provided as well as composition produced by this method.

A method of preparing dextrins, or sugars from grain that does not involve mechanical disruption of the grain prior to recovery of starch-derived comprising: treating a plant part comprising starch granules and at least one starch processing enzyme under conditions which activate the at least one enzyme thereby processing the starch granules to form an aqueous solution comprising dextrins or sugars, wherein the plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one processing enzyme; and collecting the aqueous solution comprising sugar and/or dextrins is provided.

The present invention is further directed to a method of producing fermentable sugar comprising treating a plant part comprising starch granules and at least one starch processing enzyme under conditions which activate the at least one enzyme thereby processing the starch granules to form an aqueous solution comprising dextrins or sugars, wherein the plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one processing enzyme; and

collecting the aqueous solution comprising the fermentable sugar.

Moreover, a maize plant stably transformed with a vector comprising a hyperthermophilic  $\alpha$ -amylase is provided herein. For example, preferably, a maize plant stably transformed with a vector comprising a polynucleotide sequence that encodes  $\alpha$ -amylase that is greater than 60% identical to SEQ ID NO: 1 or SEQ ID NO: 51 is encompassed.

### **Brief Description of the Figures**

Figures 1A and 1B illustrate the activity of  $\alpha$ -amylase expressed in corn kernels and in the endosperm from segregating T1 kernels from pNOV6201 plants and from six pNOV6200 lines.

Figure 2 illustrates the activity of  $\alpha$ -amylase in segregating T1 kernels from pNOV6201 lines.

Figure 3 depicts the amount of ethanol produced upon fermentation of mashes of transgenic corn containing thermostable 797GL3 alpha amylase that were subjected to liquefaction times of up to 60 minutes at 85°C and 95°C. This figure illustrates that the ethanol yield at 72 hours of fermentation was almost unchanged from 15 minutes to 60 minutes of liquefaction. Moreover, it shows that mash produced by liquefaction at 95°C produced more ethanol at each time point than mash produced by liquefaction at 85°C.

Figure 4 depicts the amount of residual starch (%) remaining after fermentation of mashes of transgenic corn containing thermostable alpha amylase that were subjected to a liquefaction time of up to 60 minutes at 85°C and 95°C. This figure illustrates that the ethanol yield at 72 hours of fermentation was almost unchanged from 15 minutes to 60 minutes of liquefaction. Moreover, it shows that mash produced by liquefaction at 95°C produced more ethanol at each time point than mash produced by liquefaction at 85°C.

Figure 5 depicts the ethanol yields for mashes of a transgenic corn, control corn, and various mixtures thereof prepared at 85°C and 95°C. This figure illustrates that the transgenic corn comprising  $\alpha$ -amylase results in significant improvement in making starch available for fermentation since there was a reduction of starch left over after fermentation.



Figure 6 depicts the amount of residual starch measured in dried stillage following fermentation for mashes of a transgenic grain, control corn, and various mixtures thereof at prepared at 85°C and 95°C.

Figure 7 depicts the ethanol yields as a function of fermentation time of a sample comprising 3% transgenic corn over a period of 20-80 hours at various pH ranges from 5.2-6.4. The figure illustrates that the fermentation conducted at a lower pH proceeds faster than at a pH of 6.0 or higher.

Figure 8 depicts the ethanol yields during fermentation of a mash comprising various weight percentages of transgenic corn from 0-12 wt% at various pH ranges from 5.2-6.4. This figure illustrates that the ethanol yield was independent of the amount of transgenic grain included in the sample.

Figure 9 shows the analysis of T2 seeds from different events transformed with pNOV 7005. High expression of pullulanase activity, compared to the non-transgenic control, can be detected in a number of events.

Figure 10A and 10B show the results of the HPLC analysis of the hydrolytic products generated by expressed pullulanase from starch in the transgenic corn flour. Incubation of the flour of pullulanase expressing corn in reaction buffer at 75 °C for 30 minutes results in production of medium chain oligosaccharides (degree of polymerization (DP) ~10-30) and short amylose chains (DP ~ 100 –200) from cornstarch. Figures 10A and 10B also show the effect of added calcium ions on the activity of the pullulanase.

Figures 11A and 11B depict the data generated from HPLC analysis of the starch hydrolysis product from two reaction mixtures. The first reaction indicated as 'Amylase' contains a mixture [1:1 (w/w)] of corn flour samples of  $\alpha$ -amylase expressing transgenic corn and non-transgenic corn A188; and the second reaction mixture 'Amylase + Pullulanase' contains a mixture [1:1 (w/w)] of corn flour samples of  $\alpha$ -amylase expressing transgenic corn and pullulanase expressing transgenic corn.

Figure 12 depicts the amount of sugar product in  $\mu\text{g}$  in 25  $\mu\text{l}$  of reaction mixture for two reaction mixtures. The first reaction indicated as 'Amylase' contains a mixture [1:1 (w/w)] of corn flour samples of  $\alpha$ -amylase expressing transgenic corn and non-transgenic corn A188; and

the second reaction mixture 'Amylase + Pullulanase' contains a mixture [1:1 (w/w)] of corn flour samples of  $\alpha$ -amylase expressing transgenic corn and pullulanase expressing transgenic corn.

Figure 13A and 13B shows the starch hydrolysis product from two sets of reaction mixtures at the end of 30 minutes incubation at 85°C and 95°C. For each set there are two reaction mixtures; the first reaction indicated as 'Amylase X Pullulanase' contains flour from transgenic corn (generated by cross pollination) expressing both the  $\alpha$ -amylase and the pullulanase, and the second reaction indicated as 'Amylase' mixture of corn flour samples of  $\alpha$ -amylase expressing transgenic corn and non-transgenic corn A188 in a ratio so as to obtain same amount of  $\alpha$ -amylase activity as is observed in the cross (Amylase X Pullulanase).

Figure 14 depicts the degradation of starch to glucose using non-transgenic corn seed (control), transgenic corn seed comprising the 797GL3  $\alpha$ -amylase, and a combination of 797GL3 transgenic corn seed with Mal A  $\alpha$ -glucosidase.

Figure 15 depicts the conversion of raw starch at room temperature or 30°C. In this figure, the reaction mixtures 1 and 2 are a combination of water and starch at room temperature and 30°C, respectively. Reaction mixtures 3 and 4 are a combination of barley  $\alpha$ -amylase and starch at room temperature and at 30°C, respectively. Reaction mixtures 5 and 6 are combinations of *Thermoanaerobacterium* glucoamylase and starch at room temperature and 30°C, respectively. Reactions mixtures 7 and 8 are combinations of barley  $\alpha$ -amylase (sigma) and *Thermoanaerobacterium* glucoamylase and starch at room temperature and 30°C, respectively. Reaction mixtures 9 and 10 are combinations of Barley alpha-amylase (sigma) control, and starch at room temperature and 30°C, respectively. The degree of polymerization (DP) of the products of the *Thermoanaerobacterium* glucoamylase is indicated.

Figure 16 depicts the production of fructose from amylase transgenic corn flour using a combination of alpha amylase, alpha glucosidase, and glucose isomerase as described in Example 19. Amylase corn flour was mixed with enzyme solutions plus water or buffer. All reactions contained 60 mg amylase flour and a total of 600 $\mu$ l of liquid and were incubated for 2 hours at 90°C.

Figure 17 depicts the peak areas of the products of reaction with 100% amylase flour from a self-processing kernel as a function of incubation time from 0-1200 minutes at 90°C.

Figure 18 depicts the peak areas of the products of reaction with 10% transgenic amylase flour from a self-processing kernel and 90% control corn flour as a function of incubation time from 0-1200 minutes at 90°C.

Figure 19 provides the results of the HPLC analysis of transgenic amylase flour incubated at 70°, 80°, 90°, or 100° C for up to 90 minutes to assess the effect of temperature on starch hydrolysis.

Figure 20 depicts ELSD peak area for samples containing 60 mg transgenic amylase flour mixed with enzyme solutions plus water or buffer under various reaction conditions. One set of reactions was buffered with 50 mM MOPS, pH 7.0 at room temperature, plus 10mM MgSO<sub>4</sub> and 1 mM CoCl<sub>2</sub>; in a second set of reactions the metal-containing buffer solution was replaced by water. All reactions were incubated for 2 hours at 90°C.

### **Detailed Description of the Invention**

In accordance with the present invention, a “self-processing” plant or plant part has incorporated therein an isolated polynucleotide encoding a processing enzyme capable of processing, e.g., modifying, starches, polysaccharides, lipids, proteins, and the like in plants, wherein the processing enzyme can be mesophilic, thermophilic or hyperthermophilic, and may be activated by grinding, addition of water, heating, or otherwise providing favorable conditions for function of the enzyme. The isolated polynucleotide encoding the processing enzyme is integrated into a plant or plant part for expression therein. Upon expression and activation of the processing enzyme, the plant or plant part of the present invention processes the substrate upon which the processing enzyme acts. Therefore, the plant or plant parts of the present invention are capable of self-processing the substrate of the enzyme upon activation of the processing enzyme contained therein in the absence of or with reduced external sources normally required for processing these substrates. As such, the transformed plants, transformed plant cells, and transformed plant parts have “built-in” processing capabilities to process desired substrates via the enzymes incorporated therein according to this invention. Preferably, the processing enzyme-encoding polynucleotide are “genetically stable,” i.e., the polynucleotide is stably maintained in the transformed plant or plant parts of the present invention and stably inherited by

progeny through successive generations.

In accordance with the present invention, methods which employ such plants and plant parts can eliminate the need to mill or otherwise physically disrupt the integrity of plant parts prior to recovery of starch-derived products. For example, the invention provides improved methods for processing corn and other grain to recover starch-derived products. The invention also provides a method which allows for the recovery of starch granules that contain levels of starch degrading enzymes, in or on the granules, that are adequate for the hydrolysis of specific bonds within the starch without the requirement for adding exogenously produced starch hydrolyzing enzymes. The invention also provides improved products from the self-processing plant or plant parts obtained by the methods of the invention.

In addition, the "self-processing" transformed plant part, e.g., grain, and transformed plant avoid major problems with existing technology, i.e., processing enzymes are typically produced by fermentation of microbes, which requires isolating the enzymes from the culture supernatants, which costs money; the isolated enzyme needs to be formulated for the particular application, and processes and machinery for adding, mixing and reacting the enzyme with its substrate must be developed. The transformed plant of the invention or a part thereof is also a source of the processing enzyme itself as well as substrates and products of that enzyme, such as sugars, amino acids, fatty acids and starch and non-starch polysaccharides. The plant of the invention may also be employed to prepare progeny plants such as hybrids and inbreds.

### **Processing Enzymes And Polynucleotides Encoding Them**

A polynucleotide encoding a processing enzyme (mesophilic, thermophilic, or hyperthermophilic) is introduced into a plant or plant part. The processing enzyme is selected based on the desired substrate upon which it acts as found in plants or transgenic plants and/or the desired end product. For example, the processing enzyme may be a starch-processing enzyme, such as a starch-degrading or starch-isomerizing enzyme, or a non-starch processing enzyme. Suitable processing enzymes include, but are not limited to, starch degrading or isomerizing enzymes including, for example,  $\alpha$ -amylase, endo or exo-1,4, or 1,6- $\alpha$ -D, glucoamylase, glucose isomerase,  $\beta$ -amylases,  $\alpha$ -glucosidases, and other exo-amylases; and

starch debranching enzymes, such as isoamylase, pullulanase, neo-pullulanase, iso-pullulanase, amylopullulanase and the like, glycosyl transferases such as cyclodextrin glycosyltransferase and the like, cellulases such as exo-1,4- $\beta$ -cellobiohydrolase, exo-1,3- $\beta$ -D-glucanase, hemicellulase,  $\beta$ -glucosidase and the like; endoglucanases such as endo-1,3- $\beta$ -glucanase and endo-1,4- $\beta$ -glucanase and the like; L-arabinases, such as endo-1,5- $\alpha$ -L-arabinase,  $\alpha$ -arabinosidases and the like; galactanases such as endo-1,4- $\beta$ -D-galactanase, endo-1,3- $\beta$ -D-galactanase,  $\beta$ -galactosidase,  $\alpha$ -galactosidase and the like; mannanases, such as endo-1,4- $\beta$ -D-mannanase,  $\beta$ -mannosidase,  $\alpha$ -mannosidase and the like; xylanases, such as endo-1,4- $\beta$ -xylanase,  $\beta$ -D-xylosidase, 1,3- $\beta$ -D-xylanase, and the like; and pectinases; and non-starch processing enzymes, including protease, glucanase, xylanase, thioredoxin/thioredoxin reductase, esterase, phytase, and lipase.

In one embodiment, the processing enzyme is a starch-degrading enzyme selected from the group of  $\alpha$ -amylase, pullulanase,  $\alpha$ -glucosidase, glucoamylase, amylopullulanase, glucose isomerase, or combinations thereof. According to this embodiment, the starch-degrading enzyme is able to allow the self-processing plant or plant part to degrade starch upon activation of the enzyme contained in the plant or plant part, as will be further described herein. The starch-degrading enzyme(s) is selected based on the desired end-products. For example, a glucose-isomerase may be selected to convert the glucose (hexose) into fructose. Alternatively, the enzyme may be selected based on the desired starch-derived end product with various chain lengths based on, e.g., a function of the extent of processing or with various branching patterns desired. For example, an  $\alpha$ -amylase, glucoamylase, or amylopullulanase can be used under short incubation times to produce dextrin products and under longer incubation times to produce shorter chain products or sugars. A pullulanase can be used to specifically hydrolyze branch points in the starch yielding a high-amylose starch, or a neopullulanase can be used to produce starch with stretches of  $\alpha$  1,4 linkages with interspersed  $\alpha$  1,6 linkages. Glucosidases could be used to produce limit dextrans, or a combination of different enzymes to make other starch derivatives.

In another embodiment, the processing enzyme is a non-starch processing enzyme selected from protease, glucanase, xylanase, phytase, lipase, cellulase, beta glucosidase and esterase. These non-starch degrading enzymes allow the self-processing plant or plant part of the

present invention to incorporate in a targeted area of the plant and, upon activation, disrupt the plant while leaving the starch granule therein intact. For example, in a preferred embodiment, the non-starch degrading enzymes target the endosperm matrix of the plant cell and, upon activation, disrupt the endosperm matrix while leaving the starch granule therein intact and more readily recoverable from the resulting material.

Combinations of processing enzymes are further envisioned by the present invention. For example, starch-processing and non-starch processing enzymes may be used in combination. Combinations of processing enzymes may be obtained by employing the use of multiple gene constructs encoding each of the enzymes. Alternatively, the individual transgenic plants stably transformed with the enzymes may be crossed by known methods to obtain a plant containing both enzymes. Another method includes the use of exogenous enzyme(s) with the transgenic plant.

The processing enzymes may be isolated or derived from any source and the polynucleotides corresponding thereto may be ascertained by one having skill in the art. For example, the processing enzyme, such as  $\alpha$ -amylase, is derived from the *Pyrococcus* (e.g., *Pyrococcus furiosus*), *Thermus*, *Thermococcus* (e.g., *Thermococcus hydrothermalis*), *Sulfolobus* (e.g., *Sulfolobus solfataricus*) *Thermotoga* (e.g., *Thermotoga maritima* and *Thermotoga neapolitana*), *Thermoanaerobacterium* (e.g. *Thermoanaerobacter tengcongensis*), *Aspergillus* (e.g., *Aspergillus shirousami* and *Aspergillus niger*), *Rhizopus* (eg., *Rhizopus oryzae*), *Thermoproteales*, *Desulfurococcus* (e.g. *Desulfurococcus amylolyticus*), *Methanobacterium thermoautotrophicum*, *Methanococcus jannaschii*, *Methanopyrus kandleri*, *Thermosynechococcus elongatus*, *Thermoplasma acidophilum*, *Thermoplasma volcanium*, *Aeropyrum pernix* and plants such as corn, barley, and rice.

The processing enzymes of the present invention are capable of being activated after being introduced and expressed in the genome of a plant. Conditions for activating the enzyme are determined for each individual enzyme and may include varying conditions such as temperature, pH, hydration, presence of metals, activating compounds, inactivating compounds, etc. For example, temperature-dependent enzymes may include mesophilic, thermophilic, and hyperthermophilic enzymes. Mesophilic enzymes typically have maximal activity at

temperatures between 20°- 65°C and are inactivated at temperatures greater than 70° C. Mesophilic enzymes have significant activity at 30 to 37°C, the activity at 30 °C is preferably at least 10% of maximal activity, more preferably at least 20% of maximal activity.

Thermophilic enzymes have a maximal activity at temperatures of between 50 and 80° C and are inactivated at temperatures greater than 80°C . A thermophilic enzyme will preferably have less than 20% of maximal activity at 30°C, more preferably less than 10% of maximal activity.

A "hyperthermophilic" enzyme has activity at even higher temperatures. Hyperthermophilic enzymes have a maximal activity at temperatures greater than 80° C and retain activity at temperatures at least 80°C, more preferably retain activity at temperatures of at least 90°C and most preferably retain activity at temperatures of at least 95°C. Hyperthermophilic enzymes also have reduced activity at low temperatures. A hyperthermophilic enzyme may have activity at 30°C that is less than 10% of maximal activity, and preferably less than 5% of maximal activity.

The polynucleotide encoding the processing enzyme is preferably modified to include codons that are optimized for expression in a selected organism such as a plant (see, e.g., Wada et al., Nucl. Acids Res., 18:2367 (1990), Murray et al., Nucl. Acids Res., 17:477 (1989), U.S. Patent Nos. 5,096,825, 5,625,136, 5,670,356 and 5,874,304). Codon optimized sequences are synthetic sequences, i.e., they do not occur in nature, and preferably encode the identical polypeptide (or an enzymatically active fragment of a full length polypeptide which has substantially the same activity as the full length polypeptide) encoded by the non-codon optimized parent polynucleotide which encodes a processing enzyme. It is preferred that the polypeptide is biochemically distinct or improved, e.g., via recursive mutagenesis of DNA encoding a particular processing enzyme, from the parent source polypeptide such that its performance in the process application is improved. Preferred polynucleotides are optimized for expression in a target host plant and encode a processing enzyme. Methods to prepare these enzymes include mutagenesis, e.g., recursive mutagenesis and selection. Methods for mutagenesis and nucleotide sequence alterations are well-known in the art. See, for example, Kunkel, Proc. Natl. Acad. Sci. USA, 82:488, (1985); Kunkel et al., Methods in Enzymol., 154:367 (1987); US

Patent No. 4,873,192; Walker and Gaastra, eds. (1983) Techniques in Molecular Biology (MacMillan Publishing Company, New York) and the references cited therein and Arnold et al., Chem. Eng. Sci., 51:5091 (1996)). Methods to optimize the expression of a nucleic acid segment in a target plant or organism are well-known in the art. Briefly, a codon usage table indicating the optimal codons used by the target organism is obtained and optimal codons are selected to replace those in the target polynucleotide and the optimized sequence is then chemically synthesized. Preferred codons for maize are described in U.S. Patent No. 5,625,136.

Complementary nucleic acids of the polynucleotides of the present invention are further envisioned. An example of low stringency conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or Northern blot is 50% formamide, e.g., hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60°C to 65°C. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C.

Moreover, polynucleotides encoding an "enzymatically active" fragment of the processing enzymes are further envisioned. As used herein, "enzymatically active" means a polypeptide fragment of the processing enzyme that has substantially the same biological activity as the processing enzyme to modify the substrate upon which the processing enzyme normally acts under appropriate conditions.

In a preferred embodiment, the polynucleotide of the present invention is a maize-optimized polynucleotide encoding  $\alpha$ -amylase, such as provided in SEQ ID NOs:2, 9, 46, and 52. In another preferred embodiment, the polynucleotide is a maize-optimized polynucleotide encoding pullulanase, such as provided in SEQ ID NOs: 4 and 25. In yet another preferred embodiment, the polynucleotide is a maize-optimized polynucleotide encoding  $\alpha$ -glucosidase as provided in SEQ ID NO:6. Another preferred polynucleotide is the maize-optimized polynucleotide encoding glucose isomerase having SEQ ID NO: 19, 21, 37, 39, 41, or 43. In



another embodiment, the maize-optimized polynucleotide encoding glucoamylase as set forth in SEQ ID NO: 46, 48, or 50 is preferred. Moreover, a maize-optimized polynucleotide for glucanase/mannanase fusion polypeptide is provided in SEQ ID NO: 57. The invention further provides for complements of such polynucleotides, which hybridize under moderate, or preferably under low stringency, hybridization conditions and which encodes a polypeptide having  $\alpha$ -amylase, pullulanase,  $\alpha$ -glucosidase, glucose isomerase, glucoamylase, glucanase, or mannanase activity, as the case may be.

The polynucleotide may be used interchangeably with "nucleic acid" or "polynucleic acid" and refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form, composed of monomers (nucleotides) containing a sugar, phosphate and a base, which is either a purine or pyrimidine. Unless specifically limited, the term encompasses nucleic acids containing known analogs of natural nucleotides, which have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues.

"Variants" or substantially similar sequences are further encompassed herein. For nucleotide sequences, variants include those sequences that, because of the degeneracy of the genetic code, encode the identical amino acid sequence of the native protein. Naturally occurring allelic variants such as these can be identified with the use of well-known molecular biology techniques, as, for example, with polymerase chain reaction (PCR), hybridization techniques, and ligation reassembly techniques. Variant nucleotide sequences also include synthetically derived nucleotide sequences, such as those generated, for example, by using site-directed mutagenesis, which encode the native protein, as well as those that encode a polypeptide having amino acid substitutions. Generally, nucleotide sequence variants of the invention will have at least 40%, 50%, 60%, preferably 70%, more preferably 80%, even more preferably 90%,

most preferably 99%, and single unit percentage identity to the native nucleotide sequence based on these classes. For example, 71%, 72%, 73% and the like, up to at least the 90% class. Variants may also include a full-length gene corresponding to an identified gene fragment.

#### **Regulatory Sequences: Promoters/Signal Sequences/Selectable Markers**

The polynucleotide sequences encoding the processing enzyme of the present invention may be operably linked to polynucleotide sequences encoding localization signals or signal sequence (at the N- or C-terminus of a polypeptide), e.g., to target the hyperthermophilic enzyme to a particular compartment within a plant. Examples of such targets include, but are not limited to, the vacuole, endoplasmic reticulum, chloroplast, amyloplast, starch granule, or cell wall, or to a particular tissue, e.g., seed. The expression of a polynucleotide encoding a processing enzyme having a signal sequence in a plant, in particular, in conjunction with the use of a tissue-specific or inducible promoter, can yield high levels of localized processing enzyme in the plant. Numerous signal sequences are known to influence the expression or targeting of a polynucleotide to a particular compartment or outside a particular compartment. Suitable signal sequences and targeting promoters are known in the art and include, but are not limited to, those provided herein.

For example, where expression in specific tissues or organs is desired, tissue-specific promoters may be used. In contrast, where gene expression in response to a stimulus is desired, inducible promoters are the regulatory elements of choice. Where continuous expression is desired throughout the cells of a plant, constitutive promoters are utilized. Additional regulatory sequences upstream and/or downstream from the core promoter sequence may be included in expression constructs of transformation vectors to bring about varying levels of expression of heterologous nucleotide sequences in a transgenic plant.

A number of plant promoters have been described with various expression characteristics. Examples of some constitutive promoters which have been described include the rice actin 1 (Wang et al., Mol. Cell. Biol., 12:3399 (1992); U.S. Patent No. 5,641,876), CaMV 35S (Odell et al., Nature, 313:810 (1985)), CaMV 19S (Lawton et al., 1987), nos (Ebert et al., 1987), Adh (Walker et al., 1987), sucrose synthase (Yang & Russell, 1990), and the ubiquitin promoters.

Vectors for use in tissue-specific targeting of genes in transgenic plants will typically include tissue-specific promoters and may also include other tissue-specific control elements such as enhancer sequences. Promoters which direct specific or enhanced expression in certain plant tissues will be known to those of skill in the art in light of the present disclosure. These include, for example, the *rbcS* promoter, specific for green tissue; the *ocs*, *nos* and *mas* promoters which have higher activity in roots or wounded leaf tissue; a truncated (-90 to +8) 35S promoter which directs enhanced expression in roots, an  $\alpha$ -tubulin gene that directs expression in roots and promoters derived from zein storage protein genes which direct expression in endosperm.

Tissue specific expression may be functionally accomplished by introducing a constitutively expressed gene (all tissues) in combination with an antisense gene that is expressed only in those tissues where the gene product is not desired. For example, a gene coding for a lipase may be introduced such that it is expressed in all tissues using the 35S promoter from Cauliflower Mosaic Virus. Expression of an antisense transcript of the lipase gene in a maize kernel, using for example a zein promoter, would prevent accumulation of the lipase protein in seed. Hence the protein encoded by the introduced gene would be present in all tissues except the kernel.

Moreover, several tissue-specific regulated genes and/or promoters have been reported in plants. Some reported tissue-specific genes include the genes encoding the seed storage proteins (such as napin, cruciferin, beta-conglycinin, and phaseolin) zein or oil body proteins (such as oleosin), or genes involved in fatty acid biosynthesis (including acyl carrier protein, stearoyl-ACP desaturase, and fatty acid desaturases (*fad 2-1*)), and other genes expressed during embryo development (such as *Bce4*, see, for example, EP 255378 and Kridl et al., Seed Science Research, 1:209 (1991)). Examples of tissue-specific promoters, which have been described include the lectin (Vodkin, Prog. Clin. Biol. Res., 138:87 (1983); Lindstrom et al., Dev. Genet., 11:160 (1990)), corn alcohol dehydrogenase 1 (Vogel et al., 1989; Dennis et al., Nucleic Acids Res., 12:3983 (1984)), corn light harvesting complex (Simpson, 1986; Bansal et al., Proc. Natl. Acad. Sci. USA, 89:3654 (1992)), corn heat shock protein (Odell et al., 1985; Rochester et al., 1986), pea small subunit RuBP carboxylase (Poulsen et al., 1986; Cashmore et al., 1983), *Ti*

plasmid mannopine synthase (Langridge et al., 1989), Ti plasmid nopaline synthase (Langridge et al., 1989), petunia chalcone isomerase (vanTunen et al., EMBO J., 7:1257(1988)), bean glycine rich protein 1 (Keller et al., Genes Dev., 3:1639 (1989)), truncated CaMV 35s (Odell et al., Nature, 313:810 (1985)), potato patatin (Wenzler et al., Plant Mol. Biol., 13:347 (1989)), root cell (Yamamoto et al., Nucleic Acids Res., 18:7449 (1990)), maize zein (Reina et al., Nucleic Acids Res., 18:6425 (1990); Kriz et al., Mol. Gen. Genet., 207:90 (1987); Wandelt et al., Nucleic Acids Res., 17:2354 (1989); Langridge et al., Cell, 34:1015 (1983); Reina et al., Nucleic Acids Res., 18:7449 (1990)), globulin-1 (Belanger et al., Genetics, 129:863 (1991)),  $\alpha$ -tubulin, cab (Sullivan et al., Mol. Gen. Genet., 215:431 (1989)), PEPCase (Hudspeth & Grula, 1989), R gene complex-associated promoters (Chandler et al., Plant Cell, 1:1175 (1989)), and chalcone synthase promoters (Franken et al., EMBO J., 10:2605 (1991)). Particularly useful for seed-specific expression is the pea vicilin promoter (Czako et al., Mol. Gen. Genet., 235:33 (1992)). (See also U.S. Pat. No. 5,625,136, herein incorporated by reference.) Other useful promoters for expression in mature leaves are those that are switched on at the onset of senescence, such as the SAG promoter from Arabidopsis (Gan et al., Science, 270:1986 (1995)).

A class of fruit-specific promoters expressed at or during anthesis through fruit development, at least until the beginning of ripening, is discussed in U.S. 4,943,674, the disclosure of which is hereby incorporated by reference. cDNA clones that are preferentially expressed in cotton fiber have been isolated (John et al., Proc. Natl. Acad. Sci. USA, 89:5769 (1992)). cDNA clones from tomato displaying differential expression during fruit development have been isolated and characterized (Mansson et al., Gen. Genet., 200:356 (1985), Slater et al., Plant Mol. Biol., 5:137 (1985)). The promoter for polygalacturonase gene is active in fruit ripening. The polygalacturonase gene is described in U.S. Patent No. 4,535,060, U.S. Patent No. 4,769,061, U.S. Patent No. 4,801,590, and U.S. Patent No. 5,107,065, which disclosures are incorporated herein by reference.

Other examples of tissue-specific promoters include those that direct expression in leaf cells following damage to the leaf (for example, from chewing insects), in tubers (for example, patatin gene promoter), and in fiber cells (an example of a developmentally-regulated fiber cell

protein is E6 (John et al., Proc. Natl. Acad. Sci. USA, 89:5769 (1992)). The E6 gene is most active in fiber, although low levels of transcripts are found in leaf, ovule and flower.

The tissue-specificity of some "tissue-specific" promoters may not be absolute and may be tested by one skilled in the art using the diphtheria toxin sequence. One can also achieve tissue-specific expression with "leaky" expression by a combination of different tissue-specific promoters (Beals et al., Plant Cell, 9:1527 (1997)). Other tissue-specific promoters can be isolated by one skilled in the art (see U.S. 5,589,379).

In one embodiment, the direction of the product from a polysaccharide hydrolysis gene, such as  $\alpha$ -amylase, may be targeted to a particular organelle such as the apoplast rather than to the cytoplasm. This is exemplified by the use of the maize  $\gamma$ -zein N-terminal signal sequence (SEQ ID NO:17), which confers apoplast-specific targeting of proteins. Directing the protein or enzyme to a specific compartment will allow the enzyme to be localized in a manner that it will not come into contact with the substrate. In this manner the enzymatic action of the enzyme will not occur until the enzyme contacts its substrate. The enzyme can be contacted with its substrate by the process of milling (physical disruption of the cell integrity), or heating the cells or plant tissues to disrupt the physical integrity of the plant cells or organs that contain the enzyme. For example a mesophilic starch-hydrolyzing enzyme can be targeted to the apoplast or to the endoplasmic reticulum and so as not to come into contact with starch granules in the amyloplast. Milling of the grain will disrupt the integrity of the grain and the starch hydrolyzing enzyme will then contact the starch granules. In this manner the potential negative effects of co-localization of an enzyme and its substrate can be circumvented.

In another embodiment, a tissue-specific promoter includes the endosperm-specific promoters such as the maize  $\gamma$ -zein promoter (exemplified by SEQ ID NO:12) or the maize ADP-gpp promoter (exemplified by SEQ ID NO:11, which includes a 5' untranslated and an intron sequence) or a Q protein promoter (exemplified by SEQ ID NO: 98) or a rice glutelin 1 promoter (exemplified in SEQ ID NO:67). Thus, the present invention includes an isolated polynucleotide comprising a promoter comprising SEQ ID NO: 11, 12, 67, or 98, a polynucleotide which hybridizes to the complement thereof under low stringency hybridization

conditions, or a fragment thereof which has promoter activity, e.g., at least 10%, and preferably at least 50%, the activity of a promoter having SEQ ID NO:11, 12, 67, or 98.

In another embodiment of the invention, the polynucleotide encodes a hyperthermophilic processing enzyme that is operably linked to a chloroplast (amyloplast) transit peptide (CTP) and a starch binding domain, e.g., from the *waxy* gene. An exemplary polynucleotide in this embodiment encodes SEQ ID NO:10 ( $\alpha$ -amylase linked to the starch binding domain from *waxy*). Other exemplary polynucleotides encode a hyperthermophilic processing enzyme linked to a signal sequence that targets the enzyme to the endoplasmic reticulum and secretion to the apoplast (exemplified by a polynucleotide encoding SEQ ID NO:13, 27, or 30, which comprises the N-terminal sequence from maize  $\gamma$ -zein operably linked to  $\alpha$ -amylase,  $\alpha$ -glucosidase, glucose isomerase, respectively), a hyperthermophilic processing enzyme linked to a signal sequence which retains the enzyme in the endoplasmic reticulum (exemplified by a polynucleotide encoding SEQ ID NO:14, 26, 28, 29, 33, 34, 35, or 36, which comprises the N-terminal sequence from maize  $\gamma$ -zein operably linked to the hyperthermophilic enzyme, which is operably linked to SEKDEL, wherein the enzyme is  $\alpha$ -amylase, malA  $\alpha$ -glucosidase, *T. maritima* glucose isomerase, *T. neapolitana* glucose isomerase), a hyperthermophilic processing enzyme linked to an N-terminal sequence that targets the enzyme to the amyloplast (exemplified by a polynucleotide encoding SEQ ID NO:15, which comprises the N-terminal amyloplast targeting sequence from *waxy* operably linked to  $\alpha$ -amylase), a hyperthermophilic fusion polypeptide which targets the enzyme to starch granules (exemplified by a polynucleotide encoding SEQ ID NO:16, which comprises the N-terminal amyloplast targeting sequence from *waxy* operably linked to an  $\alpha$ -amylase/*waxy* fusion polypeptide comprising the *waxy* starch binding domain), a hyperthermophilic processing enzyme linked to an ER retention signal (exemplified by a polynucleotide encoding SEQ ID NO:38 and 39). Moreover, a hyperthermophilic processing enzyme may be linked to a raw-starch binding site having the amino acid sequence (SEQ ID NO:53), wherein the polynucleotide encoding the processing enzyme is linked to the maize-optimized nucleic acid sequence (SEQ ID NO:54) encoding this binding site.

Several inducible promoters have been reported. Many are described in a review by Gatz, in Current Opinion in Biotechnology, 7:168 (1996) and Gatz, C., Annu. Rev. Plant Physiol.

Plant Mol. Biol., **48**:89 (1997). Examples include tetracycline repressor system, Lac repressor system, copper-inducible systems, salicylate-inducible systems (such as the PR1a system), glucocorticoid-inducible (Aoyama T. et al., N-H Plant Journal, **11**:605 (1997)) and ecdysone-inducible systems. Other inducible promoters include ABA- and turgor-inducible promoters, the promoter of the auxin-binding protein gene (Schwob et al., Plant J., **4**:423 (1993)), the UDP glucose flavonoid glycosyl-transferase gene promoter (Ralston et al., Genetics, **119**:185 (1988)), the MPI proteinase inhibitor promoter (Cordero et al., Plant J., **6**:141 (1994)), and the glyceraldehyde-3-phosphate dehydrogenase gene promoter (Kohler et al., Plant Mol. Biol., **29**:1293 (1995); Quigley et al., J. Mol. Evol., **29**:412 (1989); Martinez et al., J. Mol. Biol., **208**:551 (1989)). Also included are the benzene sulphonamide-inducible (U.S. 5364,780) and alcohol-inducible (WO 97/06269 and WO 97/06268) systems and glutathione S-transferase promoters.

Other studies have focused on genes inducibly regulated in response to environmental stress or stimuli such as increased salinity, drought, pathogen and wounding. (Graham et al., J. Biol. Chem., **260**:6555 (1985); Graham et al., J. Biol. Chem., **260**:6561 (1985), Smith et al., Planta, **168**:94 (1986)). Accumulation of metallocarboxypeptidase-inhibitor protein has been reported in leaves of wounded potato plants (Graham et al., Biochem. Biophys. Res. Comm., **101**:1164 (1981)). Other plant genes have been reported to be induced by methyl jasmonate, elicitors, heat-shock, anaerobic stress, or herbicide safeners.

Regulated expression of a chimeric transacting viral replication protein can be further regulated by other genetic strategies, such as, for example, Cre-mediated gene activation (Odell et al. Mol. Gen. Genet., **113**:369 (1990)). Thus, a DNA fragment containing 3' regulatory sequence bound by lox sites between the promoter and the replication protein coding sequence that blocks the expression of a chimeric replication gene from the promoter can be removed by Cre-mediated excision and result in the expression of the trans-acting replication gene. In this case, the chimeric Cre gene, the chimeric trans-acting replication gene, or both can be under the control of tissue- and developmental-specific or inducible promoters. An alternate genetic strategy is the use of tRNA suppressor gene. For example, the regulated expression of a tRNA suppressor gene can conditionally control expression of a trans-acting replication protein coding sequence containing an appropriate termination codon (Ulmasov et al. Plant Mol. Biol., **35**:417

(1997)). Again, either the chimeric tRNA suppressor gene, the chimeric transacting replication gene, or both can be under the control of tissue- and developmental-specific or inducible promoters.

Preferably, in the case of a multicellular organism, the promoter can also be specific to a particular tissue, organ or stage of development. Examples of such promoters include, but are not limited to, the *Zea mays* ADP-gpp and the *Zea mays*  $\gamma$ -zein promoter and the *Zea mays* globulin promoter .

Expression of a gene in a transgenic plant may be desired only in a certain time period during the development of the plant. Developmental timing is frequently correlated with tissue specific gene expression. For example, expression of zein storage proteins is initiated in the endosperm about 15 days after pollination.

Additionally, vectors may be constructed and employed in the intracellular targeting of a specific gene product within the cells of a transgenic plant or in directing a protein to the extracellular environment. This will generally be achieved by joining a DNA sequence encoding a transit or signal peptide sequence to the coding sequence of a particular gene. The resultant transit, or signal, peptide will transport the protein to a particular intracellular, or extracellular destination, respectively, and will then be post-translationally removed. Transit or signal peptides act by facilitating the transport of proteins through intracellular membranes, e.g., vacuole, vesicle, plastid and mitochondrial membranes, whereas signal peptides direct proteins through the extracellular membrane.

A signal sequence such as the maize  $\gamma$ -zein N-terminal signal sequence for targeting to the endoplasmic reticulum and secretion into the apoplast may be operably linked to a polynucleotide encoding a hyperthermophilic processing enzyme in accordance with the present invention (Torrent et al., 1997). For example, SEQ ID NOs:13, 27, and 30 provides for a polynucleotide encoding a hyperthermophilic enzyme operably linked to the N-terminal sequence from maize  $\gamma$ -zein protein. Another signal sequence is the amino acid sequence SEKDEL for retaining polypeptides in the endoplasmic reticulum (Munro and Pelham, 1987). For example, a polynucleotide encoding SEQ ID NOS:14, 26, 28, 29, 33, 34, 35, or 36, which comprises the N-terminal sequence from maize  $\gamma$ -zein operably linked to a processing enzyme



which is operably linked to SEKDEL. A polypeptide may also be targeted to the amyloplast by fusion to the waxy amyloplast targeting peptide (Klosgen et al., 1986) or to a starch granule. For example, the polynucleotide encoding a hyperthermophilic processing enzyme may be operably linked to a chloroplast (amyloplast) transit peptide (CTP) and a starch binding domain, e.g., from the *waxy* gene. SEQ ID NO:10 exemplifies  $\alpha$ -amylase linked to the starch binding domain from *waxy*. SEQ ID NO:15 exemplifies the N-terminal sequence amyloplast targeting sequence from *waxy* operably linked to  $\alpha$ -amylase. Moreover, the polynucleotide encoding the processing enzyme may be fused to target starch granules using the *waxy* starch binding domain. For example, SEQ ID NO:16 exemplifies a fusion polypeptide comprising the N-terminal amyloplast targeting sequence from *waxy* operably linked to an  $\alpha$ -amylase/*waxy* fusion polypeptide comprising the *waxy* starch binding domain.

The polynucleotides of the present invention, in addition to processing signals, may further include other regulatory sequences, as is known in the art. "Regulatory sequences" and "suitable regulatory sequences" each refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences include enhancers, promoters, translation leader sequences, introns, and polyadenylation signal sequences. They include natural and synthetic sequences as well as sequences, which may be a combination of synthetic and natural sequences.

Selectable markers may also be used in the present invention to allow for the selection of transformed plants and plant tissue, as is well-known in the art. One may desire to employ a selectable or screenable marker gene as, or in addition to, the expressible gene of interest. "Marker genes" are genes that impart a distinct phenotype to cells expressing the marker gene and thus allow such transformed cells to be distinguished from cells that do not have the marker. Such genes may encode either a selectable or screenable marker, depending on whether the marker confers a trait which one can select for by chemical means, i.e., through the use of a selective agent (e.g., a herbicide, antibiotic, or the like), or whether it is simply a trait that one can identify through observation or testing, i.e., by screening (e.g., the R-locus trait). Of course,

many examples of suitable marker genes are known to the art and can be employed in the practice of the invention.

Included within the terms selectable or screenable marker genes are also genes which encode a "secretable marker" whose secretion can be detected as a means of identifying or selecting for transformed cells. Examples include markers which encode a secretable antigen that can be identified by antibody interaction, or even secretable enzymes which can be detected by their catalytic activity. Secretable proteins fall into a number of classes, including small, diffusible proteins detectable, e.g., by ELISA; small active enzymes detectable in extracellular solution (e.g.,  $\alpha$ -amylase,  $\beta$ -lactamase, phosphinothricin acetyltransferase); and proteins that are inserted or trapped in the cell wall (e.g., proteins that include a leader sequence such as that found in the expression unit of extensin or tobacco PR-S).

With regard to selectable secretable markers, the use of a gene that encodes a protein that becomes sequestered in the cell wall, and which protein includes a unique epitope is considered to be particularly advantageous. Such a secreted antigen marker would ideally employ an epitope sequence that would provide low background in plant tissue, a promoter-leader sequence that would impart efficient expression and targeting across the plasma membrane, and would produce protein that is bound in the cell wall and yet accessible to antibodies. A normally secreted wall protein modified to include a unique epitope would satisfy all such requirements.

One example of a protein suitable for modification in this manner is extensin, or hydroxyproline rich glycoprotein (HPRG). For example, the maize HPRG (Steifel et al., The Plant Cell, 2:785 (1990)) molecule is well characterized in terms of molecular biology, expression and protein structure. However, any one of a variety of extensins and/or glycine-rich wall proteins (Keller et al., EMBO Journal, 8:1309 (1989)) could be modified by the addition of an antigenic site to create a screenable marker.

#### a. Selectable Markers

Possible selectable markers for use in connection with the present invention include, but are not limited to, a neo or nptII gene (Potrykus et al., Mol. Gen. Genet., 199:183 (1985)) which codes for kanamycin resistance and can be selected for using kanamycin, G418, and the like; a bar gene which confers resistance to the herbicide phosphinothricin; a gene which encodes an

altered EPSP synthase protein (Hinchee et al., Biotech., 6:915 (1988)) thus conferring glyphosate resistance; a nitrilase gene such as bxn from *Klebsiella ozaenae* which confers resistance to bromoxynil (Stalker et al., Science, 242:419 (1988)); a mutant acetolactate synthase gene (ALS) which confers resistance to imidazolinone, sulfonylurea or other ALS-inhibiting chemicals (European Patent Application 154,204, 1985); a methotrexate-resistant DHFR gene (Thillet et al., J. Biol. Chem., 263:12500 (1988)); a dalapon dehalogenase gene that confers resistance to the herbicide dalapon; a phosphomannose isomerase (PMI) gene; a mutated anthranilate synthase gene that confers resistance to 5-methyl tryptophan; the hph gene which confers resistance to the antibiotic hygromycin; or the mannose-6-phosphate isomerase gene (also referred to herein as the phosphomannose isomerase gene), which provides the ability to metabolize mannose (U.S. Patent Nos. 5,767,378 and 5,994,629). One skilled in the art is capable of selecting a suitable selectable marker gene for use in the present invention. Where a mutant EPSP synthase gene is employed, additional benefit may be realized through the incorporation of a suitable chloroplast transit peptide, CTP (European Patent Application 0,218,571, 1987).

An illustrative embodiment of a selectable marker gene capable of being used in systems to select transformants are the genes that encode the enzyme phosphinothricin acetyltransferase, such as the bar gene from *Streptomyces hygroscopicus* or the pat gene from *Streptomyces viridochromogenes*. The enzyme phosphinothricin acetyl transferase (PAT) inactivates the active ingredient in the herbicide bialaphos, phosphinothricin (PPT). PPT inhibits glutamine synthetase, (Murakami et al., Mol. Gen. Genet., 205:42 (1986); Twell et al., Plant Physiol., 91:1270 (1989)) causing rapid accumulation of ammonia and cell death. The success in using this selective system in conjunction with monocots was particularly surprising because of the major difficulties which have been reported in transformation of cereals (Potrykus, Trends Biotech., 7:269 (1989)).

Where one desires to employ a bialaphos resistance gene in the practice of the invention, a particularly useful gene for this purpose is the bar or pat genes obtainable from species of *Streptomyces* (e.g., ATCC No. 21,705). The cloning of the bar gene has been described (Murakami et al., Mol. Gen. Genet., 205:42 (1986); Thompson et al., EMBO Journal, 6:2519 (1987)) as has the use of the bar gene in the context of plants other than monocots (De Block et al., EMBO Journal, 6:2513 (1987); De Block et al., Plant Physiol., 91:694 (1989)).

### b. Screenable Markers

Screenable markers that may be employed include, but are not limited to, a  $\beta$ -glucuronidase or uidA gene (GUS) which encodes an enzyme for which various chromogenic substrates are known; an R-locus gene, which encodes a product that regulates the production of anthocyanin pigments (red color) in plant tissues (Dellaporta et al., in Chromosome Structure and Function, pp. 263-282 (1988)); a  $\beta$ -lactamase gene (Sutcliffe, PNAS USA, 75:3737 (1978)), which encodes an enzyme for which various chromogenic substrates are known (e.g., PADAC, a chromogenic cephalosporin); a xylE gene (Zukowsky et al., PNAS USA, 80:1101 (1983)) which encodes a catechol dioxygenase that can convert chromogenic catechols; an  $\alpha$ -amylase gene (Ikuta et al., Biotech., 8:241 (1990)); a tyrosinase gene (Katz et al., J. Gen. Microbiol., 129:2703 (1983)) which encodes an enzyme capable of oxidizing tyrosine to DOPA and dopaquinone which in turn condenses to form the easily detectable compound melanin; a  $\beta$ -galactosidase gene, which encodes an enzyme for which there are chromogenic substrates; a luciferase (lux) gene (Ow et al., Science, 234:856 (1986)), which allows for bioluminescence detection; or an aequorin gene (Prasher et al., Biochem. Biophys. Res. Comm., 126:1259 (1985)), which may be employed in calcium-sensitive bioluminescence detection, or a green fluorescent protein gene (Niedz et al., Plant Cell Reports, 14: 403 (1995)).

Genes from the maize R gene complex are contemplated to be particularly useful as screenable markers. The R gene complex in maize encodes a protein that acts to regulate the production of anthocyanin pigments in most seed and plant tissue. A gene from the R gene complex is suitable for maize transformation, because the expression of this gene in transformed cells does not harm the cells. Thus, an R gene introduced into such cells will cause the expression of a red pigment and, if stably incorporated, can be visually scored as a red sector. If a maize line carries dominant alleles for genes encoding the enzymatic intermediates in the anthocyanin biosynthetic pathway (C2, A1, A2, Bz1 and Bz2), but carries a recessive allele at the R locus, transformation of any cell from that line with R will result in red pigment formation. Exemplary lines include Wisconsin 22 which contains the rg-Stadler allele and TR112, a K55 derivative which is r-g, b, P1. Alternatively any genotype of maize can be utilized if the C1 and R alleles are introduced together. A further screenable marker contemplated for use in the

present invention is firefly luciferase, encoded by the lux gene. The presence of the lux gene in transformed cells may be detected using, for example, X-ray film, scintillation counting, fluorescent spectrophotometry, low-light video cameras, photon counting cameras or multiwell luminometry. It is also envisioned that this system may be developed for populational screening for bioluminescence, such as on tissue culture plates, or even for whole plant screening.

The polynucleotides used to transform the plant may include, but is not limited to, DNA from plant genes and non-plant genes such as those from bacteria, yeasts, animals or viruses. The introduced DNA can include modified genes, portions of genes, or chimeric genes, including genes from the same or different maize genotype. The term "chimeric gene" or "chimeric DNA" is defined as a gene or DNA sequence or segment comprising at least two DNA sequences or segments from species which do not combine DNA under natural conditions, or which DNA sequences or segments are positioned or linked in a manner which does not normally occur in the native genome of the untransformed plant.

Expression cassettes comprising the polynucleotide encoding a hyperthermophilic processing enzyme, and preferably a codon-optimized polynucleotide is further provided. It is preferred that the polynucleotide in the expression cassette (the first polynucleotide) is operably linked to regulatory sequences, such as a promoter, an enhancer, an intron, a termination sequence, or any combination thereof, and, optionally, to a second polynucleotide encoding a signal sequence (N- or C-terminal) which directs the enzyme encoded by the first polynucleotide to a particular cellular or subcellular location. Thus, a promoter and one or more signal sequences can provide for high levels of expression of the enzyme in particular locations in a plant, plant tissue or plant cell. Promoters can be constitutive promoters, inducible (conditional) promoters or tissue-specific promoters, e.g., endosperm-specific promoters such as the maize  $\gamma$ -zein promoter (exemplified by SEQ ID NO:12) or the maize ADP-gpp promoter (exemplified by SEQ ID NO:11, which includes a 5' untranslated and an intron sequence). The invention also provides an isolated polynucleotide comprising a promoter comprising SEQ ID NO:11 or 12, a polynucleotide which hybridizes to the complement thereof under low stringency hybridization conditions, or a fragment thereof which has promoter activity, e.g., at least 10%, and preferably at least 50%, the activity of a promoter having SEQ ID NO:11 or 12. Also provided are vectors which comprise the expression cassette or polynucleotide of the invention and transformed cells

comprising the polynucleotide, expression cassette or vector of the invention. A vector of the invention can comprise a polynucleotide sequence which encodes more than one hyperthermophilic processing enzyme of the invention, which sequence can be in sense or antisense orientation, and a transformed cell may comprise one or more vectors of the invention. Preferred vectors are those useful to introduce nucleic acids into plant cells.

### Transformation

The expression cassette, or a vector construct containing the expression cassette may be inserted into a cell. The expression cassette or vector construct may be carried episomally or integrated into the genome of the cell. The transformed cell may then be grown into a transgenic plant. Accordingly, the invention provides the products of the transgenic plant. Such products may include, but are not limited to, the seeds, fruit, progeny, and products of the progeny of the transgenic plant.

A variety of techniques are available and known to those skilled in the art for introduction of constructs into a cellular host. Transformation of bacteria and many eukaryotic cells may be accomplished through use of polyethylene glycol, calcium chloride, viral infection, phage infection, electroporation and other methods known in the art. Techniques for transforming plant cells or tissue include transformation with DNA employing *A. tumefaciens* or *A. rhizogenes* as the transforming agent, electroporation, DNA injection, microprojectile bombardment, particle acceleration, etc. (See, for example, EP 295959 and EP 138341).

In one embodiment, binary type vectors of Ti and Ri plasmids of *Agrobacterium* spp. Ti-derived vectors are used to transform a wide variety of higher plants, including monocotyledonous and dicotyledonous plants, such as soybean, cotton, rape, tobacco, and rice (Pacciotti et al. Bio/Technology, 3:241 (1985); Byrne et al. Plant Cell Tissue and Organ Culture, 8:3 (1987); Sukhapinda et al. Plant Mol. Biol., 8:209 (1987); Lorz et al. Mol. Gen. Genet., 199:178 (1985); Potrykus Mol. Gen. Genet., 199:183 (1985); Park et al., J. Plant Biol., 38:365 (1985); Hiei et al., Plant J., 6:271(1994)). The use of T-DNA to transform plant cells has received extensive study and is amply described (EP 120516; Hoekema, In: The Binary Plant Vector System. Offset-drukkerij Kanters B.V.; Alblaserdam (1985), Chapter V; Knauf, et al., Genetic Analysis of Host Range Expression by *Agrobacterium* In: Molecular Genetics of the

Bacteria-Plant Interaction, Puhler, A. ed., Springer-Verlag, New York, 1983, p. 245; and An. et al., EMBO J., 4:277 (1985)).

Other transformation methods are available to those skilled in the art, such as direct uptake of foreign DNA constructs (see EP 295959), techniques of electroporation (Fromm et al. Nature (London), 319:791 (1986), or high velocity ballistic bombardment with metal particles coated with the nucleic acid constructs (Kline et al. Nature (London) 327:70 (1987), and U.S. Patent No. 4,945,050). Once transformed, the cells can be regenerated by those skilled in the art. Of particular relevance are the recently described methods to transform foreign genes into commercially important crops, such as rapeseed (De Block et al., Plant Physiol. 91:694-701 (1989)), sunflower (Everett et al., Bio/Technology, 5:1201(1987)), soybean (McCabe et al., Bio/Technology, 6:923 (1988); Hinchey et al., Bio/Technology, 6:915 (1988); Chee et al., Plant Physiol., 91:1212 (1989); Christou et al., Proc. Natl. Acad. Sci USA, 86:7500 (1989) EP 301749), rice (Hiei et al., Plant J., 6:271 (1994)), and corn (Gordon Kamm et al., Plant Cell, 2:603 (1990); Fromm et al., Biotechnology, 8:833, (1990)).

Expression vectors containing genomic or synthetic fragments can be introduced into protoplasts or into intact tissues or isolated cells. Preferably expression vectors are introduced into intact tissue. General methods of culturing plant tissues are provided, for example, by Maki et al. "Procedures for Introducing Foreign DNA into Plants" in Methods in Plant Molecular Biology & Biotechnology, Glich et al. (Eds.), pp. 67-88 CRC Press (1993); and by Phillips et al. "Cell-Tissue Culture and In-Vitro Manipulation" in Corn & Corn Improvement, 3rd Edition 10, Sprague et al. (Eds.) pp. 345-387, American Society of Agronomy Inc. (1988).

In one embodiment, expression vectors may be introduced into maize or other plant tissues using a direct gene transfer method such as microprojectile-mediated delivery, DNA injection, electroporation and the like. Expression vectors are introduced into plant tissues using the microprojectile media delivery with the biolistic device. See, for example, Tomes et al. "Direct DNA transfer into intact plant cells via microprojectile bombardment" in Gamborg and Phillips (Eds.) Plant Cell, Tissue and Organ Culture: Fundamental Methods, Springer Verlag, Berlin (1995). Nevertheless, the present invention contemplates the transformation of plants with a hyperthermophilic processing enzyme in accord with known transforming methods. *Also see*, Weissinger et al., Annual Rev. Genet., 22:421 (1988); Sanford et al., Particulate Science and

Technology, 5:27 (1987) (onion); Christou et al., Plant Physiol., 87:671 (1988) (soybean); McCabe et al., Bio/Technology, 6:923 (1988) (soybean); Datta et al., Bio/Technology, 8:736 (1990) (rice); Klein et al., Proc. Natl. Acad. Sci. USA, 85:4305 (1988) (maize); Klein et al., Bio/Technology, 6:559 (1988) (maize); Klein et al., Plant Physiol., 91:440 (1988) (maize); Fromm et al., Bio/Technology, 8:833 (1990) (maize); and Gordon-Kamm et al., Plant Cell, 2, 603 (1990) (maize); Svab et al., Proc. Natl. Acad. Sci. USA, 87:8526 (1990) (tobacco chloroplast); Koziel et al., Biotechnology, 11:194 (1993) (maize); Shimamoto et al., Nature, 338:274 (1989) (rice); Christou et al., Biotechnology, 9:957 (1991) (rice); European Patent Application EP 0 332 581 (orchardgrass and other Pooideae); Vasil et al., Biotechnology, 11:1553 (1993) (wheat); Weeks et al., Plant Physiol., 102:1077 (1993) (wheat). Methods in Molecular Biology, 82. Arabidopsis Protocols Ed. Martinez-Zapater and Salinas 1998 Humana Press (Arabidopsis).

Transformation of plants can be undertaken with a single DNA molecule or multiple DNA molecules (i.e., co-transformation), and both these techniques are suitable for use with the expression cassettes and constructs of the present invention. Numerous transformation vectors are available for plant transformation, and the expression cassettes of this invention can be used in conjunction with any such vectors. The selection of vector will depend upon the preferred transformation technique and the target species for transformation.

Ultimately, the most desirable DNA segments for introduction into a monocot genome may be homologous genes or gene families which encode a desired trait (e.g., hydrolysis of proteins, lipids or polysaccharides) and which are introduced under the control of novel promoters or enhancers, etc., or perhaps even homologous or tissue specific (e.g., root-, collar/sheath-, whorl-, stalk-, earshank-, kernel- or leaf-specific) promoters or control elements. Indeed, it is envisioned that a particular use of the present invention will be the targeting of a gene in a constitutive manner or in an inducible manner.

#### Examples of Suitable Transformation Vectors

Numerous transformation vectors available for plant transformation are known to those of ordinary skill in the plant transformation arts, and the genes pertinent to this invention can be used in conjunction with any such vectors known in the art. The selection of vector will depend upon the preferred transformation technique and the target species for transformation.



#### a. Vectors Suitable for *Agrobacterium* Transformation

Many vectors are available for transformation using *Agrobacterium tumefaciens*. These typically carry at least one T-DNA border sequence and include vectors such as pBIN19 (Bevan, Nucl. Acids Res. (1984)). Below, the construction of two typical vectors suitable for *Agrobacterium* transformation is described.

##### pCIB200 and pCIB2001

The binary vectors pCIB200 and pCIB2001 are used for the construction of recombinant vectors for use with *Agrobacterium* and are constructed in the following manner. pTJS75kan is created by NarI digestion of pTJS75 (Schmidhauser & Helinski, J. Bacteriol., **164**: 446 (1985)) allowing excision of the tetracycline-resistance gene, followed by insertion of an AccI fragment from pUC4K carrying an NPTII (Messing & Vierra, Gene, **19**: 259 (1982); Bevan et al., Nature, **304**: 184 (1983); McBride et al., Plant Molecular Biology, **14**: 266 (1990)). XhoI linkers are ligated to the EcoRV fragment of PCIB7 which contains the left and right T-DNA borders, a plant selectable *nos/nptII* chimeric gene and the pUC polylinker (Rothstein et al., Gene, **53**: 153 (1987)), and the XhoI-digested fragment are cloned into SalI-digested pTJS75kan to create pCIB200 (see also EP 0 332 104, example 19). pCIB200 contains the following unique polylinker restriction sites: EcoRI, SstI, KpnI, BglII, XbaI, and SalI. pCIB2001 is a derivative of pCIB200 created by the insertion into the polylinker of additional restriction sites. Unique restriction sites in the polylinker of pCIB2001 are EcoRI, SstI, KpnI, BglII, XbaI, SalI, MluI, BclI, AvrII, ApaI, HpaI, and StuI. pCIB2001, in addition to containing these unique restriction sites also has plant and bacterial kanamycin selection, left and right T-DNA borders for *Agrobacterium*-mediated transformation, the RK2-derived *trfA* function for mobilization between *E. coli* and other hosts, and the *OriT* and *OriV* functions also from RK2. The pCIB2001 polylinker is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

pCIB10 and Hygromycin Selection Derivatives thereof:

The binary vector pCIB10 contains a gene encoding kanamycin resistance for selection in plants and T-DNA right and left border sequences and incorporates sequences from the wide host-range plasmid pRK252 allowing it to replicate in both *E. coli* and *Agrobacterium*. Its construction is described by Rothstein et al. (Gene, 53: 153 (1987)). Various derivatives of pCIB10 are constructed which incorporate the gene for hygromycin B phosphotransferase described by Gritz et al. (Gene, 25: 179 (1983)). These derivatives enable selection of transgenic plant cells on hygromycin only (pCIB743), or hygromycin and kanamycin (pCIB715, pCIB717).

b. Vectors Suitable for non-*Agrobacterium* Transformation

Transformation without the use of *Agrobacterium tumefaciens* circumvents the requirement for T-DNA sequences in the chosen transformation vector and consequently vectors lacking these sequences can be utilized in addition to vectors such as the ones described above which contain T-DNA sequences. Transformation techniques that do not rely on *Agrobacterium* include transformation via particle bombardment, protoplast uptake (e.g., PEG and electroporation) and microinjection. The choice of vector depends largely on the preferred selection for the species being transformed. Non-limiting examples of the construction of typical vectors suitable for non-*Agrobacterium* transformation is further described.

pCIB3064

pCIB3064 is a pUC-derived vector suitable for direct gene transfer techniques in combination with selection by the herbicide basta (or phosphinothricin). The plasmid pCIB246 comprises the CaMV 35S promoter in operational fusion to the *E. coli* GUS gene and the CaMV 35S transcriptional terminator and is described in the PCT published application WO 93/07278. The 35S promoter of this vector contains two ATG sequences 5' of the start site. These sites are mutated using standard PCR techniques in such a way as to remove the ATGs and generate the restriction sites SspI and PvuII. The new restriction sites are 96 and 37 bp away from the unique SalI site and 101 and 42 bp away from the actual start site. The resultant derivative of pCIB246 is designated pCIB3025. The GUS gene is then excised from pCIB3025 by digestion with SalI and SacI, the termini rendered blunt and religated to generate plasmid pCIB3060. The plasmid pJIT82 may be obtained from the John Innes Centre, Norwich and the a 400 bp *SmaI* fragment containing the *bar* gene from *Streptomyces viridochromogenes* is excised and inserted into the

*HpaI* site of pCIB3060 (Thompson et al., EMBO J, 6: 2519 (1987)). This generated pCIB3064, which comprises the *bar* gene under the control of the CaMV 35S promoter and terminator for herbicide selection, a gene for ampicillin resistance (for selection in *E. coli*) and a polylinker with the unique sites *SphI*, *PstI*, *HindIII*, and *BamHI*. This vector is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

pSOG19 and pSOG35:

The plasmid pSOG35 is a transformation vector that utilizes the *E. coli* gene dihydrofolate reductase (DHFR) as a selectable marker conferring resistance to methotrexate. PCR is used to amplify the 35S promoter (-800 bp), intron 6 from the maize *Adh1* gene (-550 bp) and 18 bp of the GUS untranslated leader sequence from pSOG10. A 250-bp fragment encoding the *E. coli* dihydrofolate reductase type II gene is also amplified by PCR and these two PCR fragments are assembled with a *SacI*-*PstI* fragment from pB1221 (Clontech) which comprises the pUC19 vector backbone and the nopaline synthase terminator. Assembly of these fragments generates pSOG19 which contains the 35S promoter in fusion with the intron 6 sequence, the GUS leader, the DHFR gene and the nopaline synthase terminator. Replacement of the GUS leader in pSOG19 with the leader sequence from Maize Chlorotic Mottle Virus (MCMV) generates the vector pSOG35. pSOG19 and pSOG35 carry the pUC gene for ampicillin resistance and have *HindIII*, *SphI*, *PstI* and *EcoRI* sites available for the cloning of foreign substances.

c. Vector Suitable for Chloroplast Transformation

For expression of a nucleotide sequence of the present invention in plant plastids, plastid transformation vector pPH143 (WO 97/32011, example 36) is used. The nucleotide sequence is inserted into pPH143 thereby replacing the PROTOX coding sequence. This vector is then used for plastid transformation and selection of transformants for spectinomycin resistance. Alternatively, the nucleotide sequence is inserted in pPH143 so that it replaces the *aadH* gene. In this case, transformants are selected for resistance to PROTOX inhibitors.

**Plant Hosts Subject to Transformation Methods**

Any plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a construct of the present invention. The term

organogenesis means a process by which shoots and roots are developed sequentially from meristematic centers while the term embryogenesis means a process by which shoots and roots develop together in a concerted fashion (not sequentially), whether from somatic cells or gametes. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include differentiated and undifferentiated tissues or plants, including but not limited to leaf disks, roots, stems, shoots, leaves, pollen, seeds, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (e.g., apical meristems, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and hypocotyl meristem), tumor tissue, and various forms of cells and culture such as single cells, protoplast, embryos, and callus tissue. The plant tissue may be in plants or in organ, tissue or cell culture.

Plants of the present invention may take a variety of forms. The plants may be chimeras of transformed cells and non-transformed cells; the plants may be clonal transformants (e.g., all cells transformed to contain the expression cassette); the plants may comprise grafts of transformed and untransformed tissues (e.g., a transformed root stock grafted to an untransformed scion in citrus species). The transformed plants may be propagated by a variety of means, such as by clonal propagation or classical breeding techniques. For example, first generation (or T1) transformed plants may be selfed to give homozygous second generation (or T2) transformed plants, and the T2 plants further propagated through classical breeding techniques. A dominant selectable marker (such as npt II) can be associated with the expression cassette to assist in breeding.

The present invention may be used for transformation of any plant species, including monocots or dicots, including, but not limited to, corn (*Zea mays*), Brassica sp. (e.g., *B. napus*, *B. rapa*, *B. juncea*), particularly those Brassica species useful as sources of seed oil, alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (e.g., pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense*, *Gossypium hirsutum*), sweet potato (*Ipomoea*

batatus), cassava (*Manihot esculenta*), coffee (*Cofea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, barley, vegetables, ornamentals, woody plants such as conifers and deciduous trees, squash, pumpkin, hemp, zucchini, apple, pear, quince, melon, plum, cherry, peach, nectarine, apricot, strawberry, grape, raspberry, blackberry, soybean, sorghum, sugarcane, rapeseed, clover, carrot, and *Arabidopsis thaliana*.

Vegetables include tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus* spp.), cauliflower, broccoli, turnip, radish, spinach, asparagus, onion, garlic, pepper, celery, and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals include azalea (*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum. Conifers that may be employed in practicing the present invention include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*), Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*). Leguminous plants include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, etc. Legumes include, but are not limited to, *Arachis*, e.g., peanuts, *Vicia*, e.g., crown vetch, hairy vetch, adzuki bean, mung bean, and chickpea, *Lupinus*, e.g., lupine, trifolium, *Phaseolus*, e.g., common bean and lima bean, *Pisum*, e.g., field bean, *Melilotus*, e.g., clover, *Medicago*, e.g., alfalfa, *Lotus*, e.g., trefoil, lens, e.g., lentil, and false indigo. Preferred forage and turf grass for use in the methods of the

invention include alfalfa, orchard grass, tall fescue, perennial ryegrass, creeping bent grass, and redtop.

Preferably, plants of the present invention include crop plants, for example, corn, alfalfa, sunflower, Brassica, soybean, cotton, safflower, peanut, sorghum, wheat, millet, tobacco, barley, rice, tomato, potato, squash, melons, legume crops, etc. Other preferred plants include Liliopsida and Panicoideae.

Once a desired DNA sequence has been transformed into a particular plant species, it may be propagated in that species or moved into other varieties of the same species, particularly including commercial varieties, using traditional breeding techniques.

Below are descriptions of representative techniques for transforming both dicotyledonous and monocotyledonous plants, as well as a representative plastid transformation technique.

a. Transformation of Dicotyledons

Transformation techniques for dicotyledons are well known in the art and include *Agrobacterium*-based techniques and techniques that do not require *Agrobacterium*. Non-*Agrobacterium* techniques involve the uptake of exogenous genetic material directly by protoplasts or cells. This can be accomplished by PEG or electroporation mediated uptake, particle bombardment-mediated delivery, or microinjection. Examples of these techniques are described by Paszkowski et al., EMBO J, **3**: 2717 (1984), Potrykus et al., Mol. Gen. Genet., **199**: 169 (1985), Reich et al., Biotechnology, **4**: 1001 (1986), and Klein et al., Nature, **327**: 70 (1987). In each case the transformed cells are regenerated to whole plants using standard techniques known in the art.

*Agrobacterium*-mediated transformation is a preferred technique for transformation of dicotyledons because of its high efficiency of transformation and its broad utility with many different species. *Agrobacterium* transformation typically involves the transfer of the binary vector carrying the foreign DNA of interest (e.g. pCIB200 or pCIB2001) to an appropriate *Agrobacterium* strain which may depend on the complement of *vir* genes carried by the host *Agrobacterium* strain either on a co-resident Ti plasmid or chromosomally (e.g., strain CIB542 for pCIB200 and pCIB2001 (Uknes et al., Plant Cell, **5**: 159 (1993))). The transfer of the recombinant binary vector to *Agrobacterium* is accomplished by a triparental mating procedure using *E. coli* carrying the recombinant binary vector, a helper *E. coli* strain which carries a

plasmid such as pRK2013 and which is able to mobilize the recombinant binary vector to the target *Agrobacterium* strain. Alternatively, the recombinant binary vector can be transferred to *Agrobacterium* by DNA transformation (Höfgen & Willmitzer, Nucl. Acids Res., 16: 9877 (1988)).

Transformation of the target plant species by recombinant *Agrobacterium* usually involves co-cultivation of the *Agrobacterium* with explants from the plant and follows protocols well known in the art. Transformed tissue is regenerated on selectable medium carrying the antibiotic or herbicide resistance marker present between the binary plasmid T-DNA borders.

The vectors may be introduced to plant cells in known ways. Preferred cells for transformation include *Agrobacterium*, monocot cells and dicots cells, including Liliopsida cells and Panicoideae cells. Preferred monocot cells are cereal cells, e.g., maize (corn), barley, and wheat, and starch accumulating dicot cells, e.g., potato.

Another approach to transforming a plant cell with a gene involves propelling inert or biologically active particles at plant tissues and cells. This technique is disclosed in U.S. Patent Nos. 4,945,050, 5,036,006, and 5,100,792. Generally, this procedure involves propelling inert or biologically active particles at the cells under conditions effective to penetrate the outer surface of the cell and afford incorporation within the interior thereof. When inert particles are utilized, the vector can be introduced into the cell by coating the particles with the vector containing the desired gene. Alternatively, the target cell can be surrounded by the vector so that the vector is carried into the cell by the wake of the particle. Biologically active particles (e.g., dried yeast cells, dried bacterium or a bacteriophage, each containing DNA sought to be introduced) can also be propelled into plant cell tissue.

#### b. Transformation of Monocotyledons

Transformation of most monocotyledon species has now also become routine. Preferred techniques include direct gene transfer into protoplasts using polyethylene glycol (PEG) or electroporation techniques, and particle bombardment into callus tissue. Transformations can be undertaken with a single DNA species or multiple DNA species (i.e., co-transformation) and both these techniques are suitable for use with this invention. Co-transformation may have the advantage of avoiding complete vector construction and of generating transgenic plants with unlinked loci for the gene of interest and the selectable marker, enabling the removal of the

selectable marker in subsequent generations, should this be regarded desirable. However, a disadvantage of the use of co-transformation is the less than 100% frequency with which separate DNA species are integrated into the genome (Schocher et al., Biotechnology, 4: 1093 1986)).

Patent Applications EP 0 292 435, EP 0 392 225, and WO 93/07278 describe techniques for the preparation of callus and protoplasts from an elite inbred line of maize, transformation of protoplasts using PEG or electroporation, and the regeneration of maize plants from transformed protoplasts. Gordon-Kamm et al. (Plant Cell, 2: 603 (1990)) and Fromm et al. (Biotechnology, 8: 833 (1990)) have published techniques for transformation of A188-derived maize line using particle bombardment. Furthermore, WO 93/07278 and Koziel et al. (Biotechnology, 11: 194 (1993)) describe techniques for the transformation of elite inbred lines of maize by particle bombardment. This technique utilizes immature maize embryos of 1.5-2.5 mm length excised from a maize ear 14-15 days after pollination and a PDS-1000He Biolistics device for bombardment.

Transformation of rice can also be undertaken by direct gene transfer techniques utilizing protoplasts or particle bombardment. Protoplast-mediated transformation has been described for *Japonica*-types and *Indica*-types (Zhang et al., Plant Cell Rep, 7: 379 (1988); Shimamoto et al., Nature, 338: 274 (1989); Datta et al., Biotechnology, 8: 736 (1990)). Both types are also routinely transformable using particle bombardment (Christou et al., Biotechnology, 9: 957 (1991)). Furthermore, WO 93/21335 describes techniques for the transformation of rice via electroporation. Patent Application EP 0 332 581 describes techniques for the generation, transformation and regeneration of Pooideae protoplasts. These techniques allow the transformation of *Dactylis* and wheat. Furthermore, wheat transformation has been described by Vasil et al. (Biotechnology, 10: 667 (1992)) using particle bombardment into cells of type C long-term regenerable callus, and also by Vasil et al. (Biotechnology, 11: 1553 (1993)) and Weeks et al. (Plant Physiol, 102: 1077 (1993)) using particle bombardment of immature embryos and immature embryo-derived callus. A preferred technique for wheat transformation, however, involves the transformation of wheat by particle bombardment of immature embryos and includes either a high sucrose or a high maltose step prior to gene delivery. Prior to bombardment, any number of embryos (0.75-1 mm in length) are plated onto MS medium with



3% sucrose (Murashiga & Skoog, Physiologia Plantarum, 15: 473 (1962)) and 3 mg/l 2,4-D for induction of somatic embryos, which is allowed to proceed in the dark. On the chosen day of bombardment, embryos are removed from the induction medium and placed onto the osmoticum (i.e., induction medium with sucrose or maltose added at the desired concentration, typically 15%). The embryos are allowed to plasmolyze for 2-3 hours and are then bombarded. Twenty embryos per target plate is typical, although not critical. An appropriate gene-carrying plasmid (such as pCIB3064 or pSG35) is precipitated onto micrometer size gold particles using standard procedures. Each plate of embryos is shot with the DuPont Biolistics® helium device using a burst pressure of about 1000 psi using a standard 80 mesh screen. After bombardment, the embryos are placed back into the dark to recover for about 24 hours (still on osmoticum). After 24 hours, the embryos are removed from the osmoticum and placed back onto induction medium where they stay for about a month before regeneration. Approximately one month later the embryo explants with developing embryogenic callus are transferred to regeneration medium (MS + 1 mg/liter NAA, 5 mg/liter GA), further containing the appropriate selection agent (10 mg/l basta in the case of pCIB3064 and 2 mg/l methotrexate in the case of pSOG35). After approximately one month, developed shoots are transferred to larger sterile containers known as "GA7s" which contain half-strength MS, 2% sucrose, and the same concentration of selection agent.

Transformation of monocotyledons using *Agrobacterium* has also been described. See, WO 94/00977 and U.S. Patent No. 5,591,616, both of which are incorporated herein by reference.

c. Transformation of Plastids

Seeds of *Nicotiana tabacum* c.v. 'Xanthi nc' are germinated seven per plate in a 1" circular array on T agar medium and bombarded 12-14 days after sowing with 1 µm tungsten particles (M10, Biorad, Hercules, CA) coated with DNA from plasmids pPH143 and pPH145 essentially as described (Svab and Maliga, PNAS, 90:913 (1993)). Bombarded seedlings are incubated on T medium for two days after which leaves are excised and placed abaxial side up in bright light (350-500 µmol photons/m<sup>2</sup>/s) on plates of RMOP medium (Svab, Hajdukiewicz and Maliga, PNAS, 87:8526 (1990)) containing 500 µg/ml spectinomycin dihydrochloride (Sigma, St. Louis, MO). Resistant shoots appearing underneath the bleached leaves three to eight weeks

after bombardment are subcloned onto the same selective medium, allowed to form callus, and secondary shoots isolated and subcloned. Complete segregation of transformed plastid genome copies (homoplasmy) in independent subclones is assessed by standard techniques of Southern blotting (Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor (1989)). BamHI/EcoRI-digested total cellular DNA (Mettler, I. J. Plant Mol Biol Reporter, 5:346 (1987)) is separated on 1% Tris-borate (TBE) agarose gels, transferred to nylon membranes (Amersham) and probed with <sup>32</sup>P-labeled random primed DNA sequences corresponding to a 0.7 kb BamHI/HindIII DNA fragment from pC8 containing a portion of the *rps7/12* plastid targeting sequence. Homoplasmic shoots are rooted aseptically on spectinomycin-containing MS/IBA medium (McBride et al., PNAS, 91:7301 (1994)) and transferred to the greenhouse.

### **Production and Characterization of Stably Transformed Plants**

Transformed plant cells are then placed in an appropriate selective medium for selection of transgenic cells, which are then grown to callus. Shoots are grown from callus and plantlets generated from the shoot by growing in rooting medium. The various constructs normally will be joined to a marker for selection in plant cells. Conveniently, the marker may be resistance to a biocide (particularly an antibiotic, such as kanamycin, G418, bleomycin, hygromycin, chloramphenicol, herbicide, or the like). The particular marker used will allow for selection of transformed cells as compared to cells lacking the DNA which has been introduced. Components of DNA constructs, including transcription/expression cassettes of this invention, may be prepared from sequences, which are native (endogenous) or foreign (exogenous) to the host. By "foreign" it is meant that the sequence is not found in the wild-type host into which the construct is introduced. Heterologous constructs will contain at least one region, which is not native to the gene from which the transcription-initiation-region is derived.

To confirm the presence of the transgenes in transgenic cells and plants, a Southern blot analysis can be performed using methods known to those skilled in the art. Integration of a polynucleic acid segment into the genome can be detected and quantitated by Southern blot, since they can be readily distinguished from constructs containing the segments through use of appropriate restriction enzymes. Expression products of the transgenes can be detected in any of

a variety of ways, depending upon the nature of the product, and include Western blot and enzyme assay. One particularly useful way to quantitate protein expression and to detect replication in different plant tissues is to use a reporter gene, such as GUS. Once transgenic plants have been obtained, they may be grown to produce plant tissues or parts having the desired phenotype. The plant tissue or plant parts may be harvested, and/or the seed collected. The seed may serve as a source for growing additional plants with tissues or parts having the desired characteristics.

The invention thus provides a transformed plant or plant part, such as an ear, seed, fruit, grain, stover, chaff, or bagasse comprising at least one polynucleotide, expression cassette or vector of the invention, methods of making such a plant and methods of using such a plant or a part thereof. The transformed plant or plant part expresses a processing enzyme, optionally localized in a particular cellular or subcellular compartment of a certain tissue or in developing grain. For instance, the invention provides a transformed plant part comprising at least one starch processing enzyme present in the cells of the plant, wherein the plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one starch processing enzyme. The processing enzyme does not act on the target substrate unless activated by methods such as heating, grinding, or other methods, which allow the enzyme to contact the substrate under conditions where the enzyme is active

### **Exemplary Methods of the Present Invention**

The self-processing plants and plant parts of the present invention may be used in various methods employing the processing enzymes (mesophilic, thermophilic, or hyperthermophilic) expressed and activated therein. In accordance with the present invention, a transgenic plant part obtained from a transgenic plant the genome of which is augmented with at least one processing enzyme, is placed under conditions in which the processing enzyme is expressed and activated. Upon activation, the processing enzyme is activated and functions to act on the substrate in which it normally acts to obtain the desired result. For example, the starch-processing enzymes act upon starch to degrade, hydrolyze, isomerize, or otherwise modify to obtain the desired result upon activation. Non-starch processing enzymes may be used to disrupt the plant cell membrane in order to facilitate the extraction of starch, lipids, amino acids, or other products

from the plants. Moreover, non-hyperthermophilic and hyperthermophilic enzymes may be used in combination in the self-processing plant or plant parts of the present invention. For example, a mesophilic non-starch degrading enzyme may be activated to disrupt the plant cell membrane for starch extraction, and subsequently, a hyperthermophilic starch-degrading enzyme may then be activated in the self-processing plant to degrade the starch.

Enzymes expressed in grain can be activated by placing the plant or plant part containing them in conditions in which their activity is promoted. For example, one or more of the following techniques may be used: The plant part may be contacted with water, which provides a substrate for a hydrolytic enzyme and thus will activate the enzyme. The plant part may be contacted with water which will allow enzyme to migrate from the compartment into which it was deposited during development of the plant part and thus to associate with its substrate. Movement of the enzyme is possible because compartmentalization is breached during maturation, drying of grain and re-hydration. The intact or cracked grain may be contacted with water which will allow enzyme to migrate from the compartment into which it was deposited during development of the plant part and thus to associate with its substrate. Enzymes can also be activated by addition of an activating compound. For example, a calcium-dependent enzyme can be activated by addition of calcium. Other activating compounds may be determined by those skilled in the art. Enzymes can be activated by removal of an inactivator. For example, there are known peptide inhibitors of amylase enzymes, the amylase could be co-expressed with an amylase inhibitor and then activated by addition of a protease. Enzymes can be activated by alteration of pH to one at which the enzyme is most active. Enzymes can also be activated by increasing temperature. An enzyme generally increases in activity up to the maximal temperature for that enzyme. A mesophilic enzyme will increase in activity from the level of activity ambient temperature up to the temperature at which it loses activity which is typically less than or equal to 70 °C. Similarly thermophilic and hyperthermophilic enzymes can also be activated by increasing temperature. Thermophilic enzymes can be activated by heating to temperatures up to the maximal temperature of activity or of stability. For a thermophilic enzyme the maximal temperatures of stability and activity will generally be between 70 and 85 °C. Hyperthermophilic enzymes will have the even greater relative activation than mesophilic or

thermophilic enzymes because of the greater potential change in temperature from 25 °C up to 85 °C to 95 °C or even 100 °C. The increased temperature may be achieved by any method, for example by heating such as by baking, boiling, heating, steaming, electrical discharge or any combination thereof. Moreover, in plants expressing mesophilic or thermophilic enzyme(s), activation of the enzyme may be accomplished by grinding, thereby allowing the enzyme to contact the substrate.

The optimal conditions, e.g., temperature, hydration, pH, etc, may be determined by one having skill in the art and may depend upon the individual enzyme being employed and the desired application of the enzyme.

The present invention further provides for the use of exogenous enzymes that may assist in a particular process. For example, the use of a self-processing plant or plant part of the present invention may be used in combination with an exogenously provided enzyme to facilitate the reaction. As an example, transgenic  $\alpha$ -amylase corn may be used in combination with other starch-processing enzymes, such as pullulanase,  $\alpha$ -glucosidase, glucose isomerase, mannanases, hemicellulases, etc., to hydrolyze starch or produce ethanol. In fact, it has been found that combinations of the transgenic  $\alpha$ -amylase corn with such enzymes has unexpectedly provided superior degrees of starch conversion relative to the use of transgenic  $\alpha$ -amylase corn alone.

Example of suitable methods contemplated herein are provided.

a. Starch Extraction From Plants

The invention provides for a method of facilitating the extraction of starch from plants. In particular, at least one polynucleotide encoding a processing enzyme that disrupts the physically restraining matrix of the endosperm (cell walls, non-starch polysaccharide, and protein matrix) is introduced to a plant so that the enzyme is preferably in close physical proximity to starch granules in the plant. In this embodiment of the invention, transformed plants express one or more protease, glucanase, xylanase, thioredoxin/thioredoxin reductase, cellulase, phytase, lipase, beta glucosidase, esterase and the like, but not enzymes that have any starch degrading activity, so as to maintain the integrity of the starch granules. The expression of these enzymes in a plant part such as grain thus improves the process characteristics of grain. The processing enzyme may be mesophilic, thermophilic, or hyperthermophilic. In one example,

grain from a transformed plant of the invention is heat dried, likely inactivating non-hyperthermophilic processing enzymes and improving seed integrity. Grain (or cracked grain) is steeped at low temperatures or high temperatures (where time is of the essence) with high or low moisture content or conditions (see Primary Cereal Processing, Gordon and Willm, eds., pp. 319-337 (1994), the disclosure of which is incorporated herein), with or without sulphur dioxide. Upon reaching elevated temperatures, optionally at certain moisture conditions, the integrity of the endosperm matrix is disrupted by activating the enzymes, e.g., proteases, xylanases, phytase or glucanases which degrade the proteins and non-starch polysaccharides present in the endosperm leaving the starch granule therein intact and more readily recoverable from the resulting material. Further, the proteins and non-starch polysaccharides in the effluent are at least partially degraded and highly concentrated, and so may be used for improved animal feed, food, or as media components for the fermentation of microorganisms. The effluent is considered a corn-steep liquor with improved composition.

Thus, the invention provides a method to prepare starch granules. The method comprises treating grain, for example cracked grain, which comprises at least one non-starch processing enzyme under conditions which activate the at least one enzyme, yielding a mixture comprising starch granules and non-starch degradation products, e.g., digested endosperm matrix products. The non-starch processing enzyme may be mesophilic, thermophilic, or hyperthermophilic. After activation of the enzyme, the starch granules are separated from the mixture. The grain is obtained from a transformed plant, the genome of which comprises (is augmented with) an expression cassette encoding the at least one processing enzyme. For example, the processing enzyme may be a protease, glucanase, xylanase, phytase, thioredoxin/thioredoxin reductase, esterase cellulase, lipase, or a beta glucosidase. The processing enzyme may be hyperthermophilic. The grain can be treated under low or high moisture conditions, in the presence or absence of sulfur dioxide. Depending on the activity and expression level of the processing enzyme in the grain from the transgenic plant, the transgenic grain may be mixed with commodity grain prior to or during processing. Also provided are products obtained by the method such as starch, non-starch products and improved steepwater comprising at least one additional component.

b. Starch-Processing Methods

Transformed plants or plant parts of the present invention may comprise starch-degrading enzymes as disclosed herein that degrade starch granules to dextrins, other modified starches, or hexoses (*e.g.*,  $\alpha$ -amylase, pullulanase,  $\alpha$ -glucosidase, glucoamylase, amylopullulanase) or convert glucose into fructose (*e.g.*, glucose isomerase). Preferably, the starch-degrading enzyme is selected from  $\alpha$ -amylase,  $\alpha$ -glucosidase, glucoamylase, pullulanase, neopullulanase, amylopullulanase, glucose isomerase, and combinations thereof is used to transform the grain. Moreover, preferably, the enzyme is operably linked to a promoter and to a signal sequence that targets the enzyme to the starch granule, an amyloplast, the apoplast, or the endoplasmic reticulum. Most preferably, the enzyme is expressed in the endosperm, and particularly, corn endosperm, and localized to one or more cellular compartments, or within the starch granule itself. The preferred plant part is grain. Preferred plant parts are those from corn, wheat, barley, rye, oat, sugar cane, or rice.

In accordance with one starch-degrading method of the present invention, the transformed grain accumulates the starch-degrading enzyme in starch granules, is steeped at conventional temperatures of 50°C-60°C, and wet-milled as is known in the art. Preferably, the starch-degrading enzyme is hyperthermophilic. Because of sub-cellular targeting of the enzyme to the starch granule, or by virtue of the association of the enzyme with the starch granule, by contacting the enzyme and starch granule during the wet-milling process at the conventional temperatures, the processing enzyme is co-purified with the starch granules to obtain the starch granules/enzyme mixture. Subsequent to the recovery of the starch granules/enzyme mixture, the enzyme is then activated by providing favorable conditions for the activity of the enzyme. For example, the processing may be performed in various conditions of moisture and/or temperature to facilitate the partial (in order to make derivatized starches or dextrins) or complete hydrolysis of the starch into hexoses. Syrups containing high dextrose or fructose equivalents are obtained in this manner. This method effectively reduces the time, energy, and enzyme costs and the efficiency with which starch is converted to the corresponding hexose, and the efficiency of the production of products, like high sugar steepwater and higher dextrose equivalent syrups, are increased.

In another embodiment, a plant, or a product of the plant such as a fruit or grain, or flour made from the grain that expresses the enzyme is treated to activate the enzyme and convert polysaccharides expressed and contained within the plant into sugars. Preferably, the enzyme is fused to a signal sequence that targets the enzyme to a starch granule, an amyloplast, the apoplast or to the endoplasmic reticulum as disclosed herein. The sugar produced may then be isolated or recovered from the plant or the product of the plant. In another embodiment, a processing enzyme able to convert polysaccharides into sugars is placed under the control of an inducible promoter according to methods known in the art and disclosed herein. The processing enzyme may be mesophilic, thermophilic or hyperthermophilic. The plant is grown to a desired stage and the promoter is induced causing expression of the enzyme and conversion of the polysaccharides, within the plant or product of the plant, to sugars. Preferably the enzyme is operably linked to a signal sequence that targets the enzyme to a starch granule, an amyloplast, an apoplast or to the endoplasmic reticulum. In another embodiment, a transformed plant is produced that expresses a processing enzyme able to convert starch into sugar. The enzyme is fused to a signal sequence that targets the enzyme to a starch granule within the plant. Starch is then isolated from the transformed plant that contains the enzyme expressed by the transformed plant. The enzyme contained in the isolated starch may then be activated to convert the starch into sugar. The enzyme may be mesophilic, thermophilic, or hyperthermophilic. Examples of hyperthermophilic enzymes able to convert starch to sugar are provided herein. The methods may be used with any plant which produces a polysaccharide and that can express an enzyme able to convert a polysaccharide into sugars or hydrolyzed starch product such as dextrin, maltooligosaccharide, glucose and/or mixtures thereof.

The invention provides a method to produce dextrans and altered starches from a plant, or a product from a plant, that has been transformed with a processing enzyme which hydrolyses certain covalent bonds of a polysaccharide to form a polysaccharide derivative. In one embodiment, a plant, or a product of the plant such as a fruit or grain, or flour made from the grain that expresses the enzyme is placed under conditions sufficient to activate the enzyme and convert polysaccharides contained within the plant into polysaccharides of reduced molecular weight. Preferably, the enzyme is fused to a signal sequence that targets the enzyme to a starch granule, an amyloplast, the apoplast or to the endoplasmic reticulum as disclosed herein. The



dextrin or derivative starch produced may then be isolated or recovered from the plant or the product of the plant. In another embodiment, a processing enzyme able to convert polysaccharides into dextrans or altered starches is placed under the control of an inducible promoter according to methods known in the art and disclosed herein. The plant is grown to a desired stage and the promoter is induced causing expression of the enzyme and conversion of the polysaccharides, within the plant or product of the plant, to dextrans or altered starches. Preferably the enzyme is  $\alpha$ -amylase, pullulanase, iso or neo-pullulanase and is operably linked to a signal sequence that targets the enzyme to a starch granule, an amyloplast, the apoplast or to the endoplasmic reticulum. In one embodiment, the enzyme is targeted to the apoplast or to the endoplasmic reticulum. In yet another embodiment, a transformed plant is produced that expresses an enzyme able to convert starch into dextrans or altered starches. The enzyme is fused to a signal sequence that targets the enzyme to a starch granule within the plant. Starch is then isolated from the transformed plant that contains the enzyme expressed by the transformed plant. The enzyme contained in the isolated starch may then be activated under conditions sufficient for activation to convert the starch into dextrans or altered starches. Examples of hyperthermophilic enzymes, for example, able to convert starch to hydrolyzed starch products are provided herein. The methods may be used with any plant which produces a polysaccharide and that can express an enzyme able to convert a polysaccharide into sugar.

In another embodiment, grain from transformed plants of the invention that accumulate starch-degrading enzymes that degrade linkages in starch granules to dextrans, modified starches or hexose (*e.g.*,  $\alpha$ -amylase, pullulanase,  $\alpha$ -glucosidase, glucoamylase, amylopullulanase) is steeped under conditions favoring the activity of the starch degrading enzyme for various periods of time. The resulting mixture may contain high levels of the starch-derived product. The use of such grain: 1) eliminates the need to mill the grain, or otherwise process the grain to first obtain starch granules, 2) makes the starch more accessible to enzymes by virtue of placing the enzymes directly within the endosperm tissue of the grain, and 3) eliminates the need for microbially produced starch-hydrolyzing enzymes. Thus, the entire process of wet-milling prior to hexose recovery is eliminated by simply heating grain, preferably corn grain, in the presence of water to allow the enzymes to act on the starch.

This process can also be employed for the production of ethanol, high fructose syrups, hexose (glucose) containing fermentation media, or any other use of starch that does not require the refinement of grain components.

The invention further provides a method of preparing dextrin, maltooligosaccharides, and/or sugar involving treating a plant part comprising starch granules and at least one starch processing enzyme under conditions so as to activate the at least one enzyme thereby digesting starch granules to form an aqueous solution comprising sugars. The plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one processing enzyme. The aqueous solution comprising dextrans, maltooligosaccharides, and/or sugar is then collected. In one embodiment, the processing enzyme is  $\alpha$ -amylase,  $\alpha$ -glucosidase, pullulanase, glucoamylase, amylopullulanase, glucose isomerase, or any combination thereof. Preferably, the enzyme is hyperthermophilic. In another embodiment, the method further comprises isolating the dextrans, maltooligosaccharides, and/or sugar.

c. Improved Corn Varieties

The invention also provides for the production of improved corn varieties (and varieties of other crops) that have normal levels of starch accumulation, and accumulate sufficient levels of amylolytic enzyme(s) in their endosperm, or starch accumulating organ, such that upon activation of the enzyme contained therein, such as by boiling or heating the plant or a part thereof in the case of a hyperthermophilic enzyme, the enzyme(s) is activated and facilitates the rapid conversion of the starch into simple sugars. These simple sugars (primarily glucose) will provide sweetness to the treated corn. The resulting corn plant is an improved variety for dual use as a grain

producing hybrid and as sweet corn. Thus, the invention provides a method to produce hyper-sweet corn, comprising treating transformed corn or a part thereof, the genome of which is augmented with and expresses in endosperm an expression cassette comprising a promoter operably linked to a first polynucleotide encoding at least one amylolytic enzyme, conditions which activate the at least one enzyme so as to convert polysaccharides in the corn into sugar, yielding hypersweet corn. The promoter may be a constitutive promoter, a seed-

specific promoter, or an endosperm-specific promoter which is linked to a polynucleotide sequence which encodes a processing enzyme such as  $\alpha$ -amylase, e.g., one comprising SEQ ID NO: 13, 14, or 16. Preferably, the enzyme is hyperthermophilic. In one embodiment, the expression cassette further comprises a second polynucleotide which encodes a signal sequence operably linked to the enzyme encoded by the first polynucleotide. Exemplary signal sequences in this embodiment of the invention direct the enzyme to apoplast, the endoplasmic reticulum, a starch granule, or to an amyloplast. The corn plant is grown such that the ears with kernels are formed and then the promoter is induced to cause the enzyme to be expressed and convert polysaccharide contained within the plant into sugar.

d. Self-Fermenting Plants

In another embodiment of the invention, plants, such as corn, rice, wheat, or sugar cane are engineered to accumulate large quantities of processing enzymes in their cell walls, e.g., xylanases, cellulases, hemicellulases, glucanases, pectinases, lipases, esterases, beta glucosidases, phytases, proteases and the like (non-starch polysaccharide degrading enzymes). Following the harvesting of the grain component (or sugar in the case of sugar cane), the stover, chaff, or bagasse is used as a source of the enzyme, which was targeted for expression and accumulation in the cell walls, and as a source of biomass. The stover (or other left-over tissue) is used as a feedstock in a process to recover fermentable sugars. The process of obtaining the fermentable sugars consists of activating the non-starch polysaccharide degrading enzyme. For example, activation may comprise heating the plant tissue in the presence of water for periods of time adequate for the hydrolysis of the non-starch polysaccharide into the resulting sugars. Thus, this self-processing stover produces the enzymes required for conversion of polysaccharides into monosaccharides, essentially at no incremental cost as they are a component of the feedstock. Further, the temperature-dependent enzymes have no detrimental effects on plant growth and development, and cell wall targeting, even targeting into polysaccharide microfibrils by virtue of cellulose/xylose binding domains fused to the protein, improves the accessibility of the substrate to the enzyme.

Thus, the invention also provides a method of using a transformed plant part comprising at least one non-starch polysaccharide processing enzyme in the cell wall of the cells of the plant

part. The method comprises treating a transformed plant part comprising at least one non-starch polysaccharide processing enzyme under conditions which activate the at least one enzyme thereby digesting starch granules to form an aqueous solution comprising sugars, wherein the plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one non-starch polysaccharide processing enzyme; and collecting the aqueous solution comprising the sugars. The invention also includes a transformed plant or plant part comprising at least one non-starch polysaccharide processing enzyme present in the cell or cell wall of the cells of the plant or plant part. The plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one non-starch processing enzyme, e.g., a xylanase, cellulase, glucanase, pectinase, lipase, esterase, beta glucosidase, phytase, protease or any combination thereof.

e. Aqueous Phase High In Protein and Sugar Content

In yet another embodiment, proteases and lipases are engineered to accumulate in seeds, e.g., soybean seeds. After activation of the protease or lipase, such as, for example, by heating, these enzymes in the seeds hydrolyze the lipid and storage proteins present in soybeans during processing. Soluble products comprising amino acids, which can be used as feed, food or fermentation media, and fatty acids, can thus be obtained. Polysaccharides are typically found in the insoluble fraction of processed grain. However, by combining polysaccharide degrading enzyme expression and accumulation in seeds, proteins and polysaccharides can be hydrolyzed and are found in the aqueous phase. For example, zeins from corn and storage protein and non-starch polysaccharides from soybean can be solubilized in this manner. Components of the aqueous and hydrophobic phases can be easily separated by extraction with organic solvent or supercritical carbon dioxide. Thus, what is provided is a method for producing an aqueous extract of grain that contains higher levels of protein, amino acids, sugars or saccharides.

f. Self-Processing Fermentation

The invention provides a method to produce ethanol, a fermented beverage, or other fermentation-derived product(s). The method involves obtaining a plant, or the product or part of a plant, or plant derivative such as grain flour, wherein a processing enzyme that converts polysaccharides into sugar is expressed. The plant, or product thereof, is treated such that sugar is produced by conversion of the polysaccharide as described above. The sugars and other components of the plant are then fermented to form ethanol or a fermented beverage, or other fermentation-derived products, according to methods known in the art. See, for example, U.S. Patent No.: 4,929,452. Briefly the sugar produced by conversion of polysaccharides is incubated with yeast under conditions that promote conversion of the sugar into ethanol. A suitable yeast includes high alcohol-tolerant and high-sugar tolerant strains of yeast, such as, for example, the yeast, *S. cerevisiae* ATCC No. 20867. This strain was deposited with the American Type Culture Collection, Rockville, MD, on Sept. 17, 1987 and assigned ATCC No. 20867. The fermented product or fermented beverage may then be distilled to isolate ethanol or a distilled beverage, or the fermentation product otherwise recovered. The plant used in this method may be any plant that contains a polysaccharide and is able to express an enzyme of the invention. Many such plants are disclosed herein. Preferably the plant is one that is grown commercially. More

preferably the plant is one that is normally used to produce ethanol or fermented beverages, or fermented products, such as, for example, wheat, barley, corn, rye, potato, grapes or rice.

The method comprises treating a plant part comprising at least one polysaccharide processing enzyme under conditions to activate the at least one enzyme thereby digesting polysaccharide in the plant part to form fermentable sugar. The polysaccharide processing enzyme may be mesophilic, thermophilic, or hyperthermophilic. The plant part is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one polysaccharide processing enzyme. Plant parts for this embodiment of the invention include, but are not limited to, grain, fruit, seed, stalk, wood, vegetable or root. Plants include but are not limited to oat, barley, wheat, berry, grape, rye, corn, rice, potato, sugar beet, sugar cane, pineapple, grass and tree. The plant part may be combined with commodity grain or other commercially available substrates; the source of the substrate for processing may be a source other than the self-processing plant. The fermentable sugar is then incubated under conditions that promote the conversion of the fermentable sugar into ethanol, e.g., with yeast and/or other microbes. In an embodiment, the plant part is derived from corn transformed with  $\alpha$ -amylase, which has been found to reduce the amount of time and cost of fermentation.

It has been found that the amount of residual starch is reduced when transgenic corn made in accordance with the present invention expressing a thermostable  $\alpha$ -amylase, for example, is used in fermentation. This indicates that more starch is solubilized during fermentation. The reduced amount of residual starch results in the distillers' grains having higher protein content by weight and higher value. Moreover, the fermentation of the transgenic corn of the present invention allows the liquefaction process to be performed at a lower pH, resulting in savings in the cost of chemicals used to adjust the pH, at a higher temperature, e.g., greater than 85°C, preferably, greater than 90°C, more preferably, 95°C or higher, resulting in shorter liquefaction times and more complete solubilization of starch, and reduction of liquefaction times, all resulting in efficient fermentation reactions with higher yields of ethanol.

Moreover, it has been found that contacting conventional plant parts with even a small portion of the transgenic plant made in accordance with the present invention may reduce the fermentation time and costs associated therewith. As such, the present invention relates to the reduction in the fermentation time for plants comprising subjecting a transgenic plant part from a

plant comprising a polysaccharide processing enzyme that converts polysaccharides into sugar relative to the use of a plant part not comprising the polysaccharide processing enzyme.

g. Raw Starch Processing Enzymes And Polynucleotides Encoding Them

A polynucleotide encoding a mesophilic processing enzyme(s) is introduced into a plant or plant part. In an embodiment, the polynucleotide of the present invention is a maize-optimized polynucleotide such as provided in SEQ ID NOs: 48, 50, and 59, encoding a glucoamylase, such as provided in SEQ ID NOs: 47, and 49. In another embodiment, the polynucleotide of the present invention is a maize-optimized polynucleotide such as provided in SEQ ID NO: 52, encoding an alpha-amylase, such as provided in SEQ ID NO: 51. Moreover, fusion products of processing enzymes are further contemplated. In one embodiment, the polynucleotide of the present invention is a maize-optimized polynucleotide such as provided in SEQ ID NO: 46, encoding an alpha-amylase and glucoamylase fusion, such as provided in SEQ ID NO: 45. Combinations of processing enzymes are further envisioned by the present invention. For example, a combination of starch-processing enzymes and non-starch processing enzymes is contemplated herein. Such combinations of processing enzymes may be obtained by employing the use of multiple gene constructs encoding each of the enzymes. Alternatively, the individual transgenic plants stably transformed with the enzymes may be crossed by known methods to obtain a plant containing both enzymes. Another method includes the use of exogenous enzyme(s) with the transgenic plant.

The source of the starch-processing and non-starch processing enzymes may be isolated or derived from any source and the polynucleotides corresponding thereto may be ascertained by one having skill in the art. The  $\alpha$ -amylase may be derived from *Aspergillus* (e.g., *Aspergillus shirousami* and *Aspergillus niger*), *Rhizopus* (eg., *Rhizopus oryzae*), and plants such as corn, barley, and rice. The glucoamylase may be derived from *Aspergillus* (e.g., *Aspergillus shirousami* and *Aspergillus niger*), *Rhizopus* (eg., *Rhizopus oryzae*), and *Thermoanaerobacter* (eg., *Thermoanaerobacter thermosaccharolyticum*).

In another embodiment of the invention, the polynucleotide encodes a mesophilic starch-processing enzyme that is operably linked to a maize-optimized polynucleotide such as provided in SEQ ID NO: 54, encoding a raw starch binding domain, such as provided in SEQ ID NO: 53.

In another embodiment, a tissue-specific promoter includes the endosperm-specific promoters such as the maize  $\gamma$ -zein promoter (exemplified by SEQ ID NO:12) or the maize ADP-gpp promoter (exemplified by SEQ ID NO:11, which includes a 5' untranslated and an intron sequence) or a Q protein promoter (exemplified by SEQ ID NO: 98) or a rice glutelin promoter (exemplified by SEQ ID NO: 67) . Thus, the present invention includes an isolated polynucleotide comprising a promoter comprising SEQ ID NO: 11, 12, 67, or 98, a polynucleotide which hybridizes to the complement thereof under low stringency hybridization conditions, or a fragment thereof which has promoter activity, e.g., at least 10%, and preferably at least 50%, the activity of a promoter having SEQ ID NO:11, 12, 67 or 98.

In one embodiment, the product from a starch-hydrolysis gene, such as  $\alpha$ -amylase, glucoamylase, or  $\alpha$ -amylase/glucoamylase fusion may be targeted to a particular organelle or location such as the endoplasmic reticulum or apoplast, rather than to the cytoplasm. This is exemplified by the use of the maize  $\gamma$ -zein N-terminal signal sequence (SEQ ID NO:17), which confers apoplast-specific targeting of proteins, and the use of the  $\gamma$ -zein N-terminal signal sequence (SEQ ID NO:17) which is operably linked to the processing enzyme that is operably linked to the sequence SEKDEL for retention in the endoplasmic reticulum. Directing the protein or enzyme to a specific compartment will allow the enzyme to be localized in a manner that it will not come into contact with the substrate. In this manner the enzymatic action of the enzyme will not occur until the enzyme contacts its substrate. The enzyme can be contacted with its substrate by the process of milling (physical disruption of the cell integrity) and hydrating. For example, a mesophilic starch-hydrolyzing enzyme can be targeted to the apoplast or to the endoplasmic reticulum and will therefore not come into contact with starch granules in the amyloplast. Milling of the grain will disrupt the integrity of the grain and the starch hydrolyzing enzyme will then contact the starch granules. In this manner the potential negative effects of co-localization of an enzyme and its substrate can be circumvented.

h. Food Products Without Added Sweetener

Also provided is a method to produce a sweetened farinaceous food product without adding additional sweetener. Examples of farinaceous products include, but are not limited to, breakfast food, ready to eat food, baked food, pasta and cereal products such as



breakfast cereal. The method comprises treating a plant part comprising at least one starch processing enzyme under conditions which activate the starch processing enzyme, thereby processing starch granules in the plant part to sugars so as to form a sweetened product, e.g., relative to the product produced by processing starch granules from a plant part which does not comprise the hyperthermophilic enzyme. Preferably, the starch processing enzyme is hyperthermophilic and is activated by heating, such as by baking, boiling, heating, steaming, electrical discharge, or any combination thereof. The plant part is obtained from a transformed plant, for instance from transformed soybean, rye, oat, barley, wheat, corn, rice or sugar cane, the genome of which is augmented with an expression cassette encoding the at least one hyperthermophilic starch processing enzyme, e.g.,  $\alpha$ -amylase,  $\alpha$ -glucosidase, glucoamylase, pullulanase, glucose isomerase, or any combination thereof. The sweetened product is then processed into a farinaceous food product. The invention also provides a farinaceous food product, e.g., a cereal food, a breakfast food, a ready to eat food, or a baked food, produced by the method. The farinaceous food product may be formed from the sweetened product and water, and may contain malt, flavorings, vitamins, minerals, coloring agents or any combination thereof.

The enzyme may be activated to convert polysaccharides contained within the plant material into sugar prior to inclusion of the plant material into the cereal product or during the processing of the cereal product. Accordingly, polysaccharides contained within the plant material may be converted into sugar by activating the material, such as by heating in the case of a hyperthermophilic enzyme, prior to inclusion in the farinaceous product. The plant material containing sugar produced by conversion of the polysaccharides is then added to the product to produce a sweetened product. Alternatively, the polysaccharides may be converted into sugars by the enzyme during the processing of the farinaceous product. Examples of processes used to make cereal products are well known in the art and include heating, baking, boiling and the like as described in U.S. Patent Nos.: 6,183,788; 6,159,530; 6,149,965; 4,988,521 and 5,368,870.

Briefly, dough may be prepared by blending various dry ingredients together with water and cooking to gelatinize the starchy components and to develop a cooked flavor. The cooked material can then be mechanically worked to form a cooked dough, such as cereal dough. The

dry ingredients may include various additives such as sugars, starch, salt, vitamins, minerals, colorings, flavorings, salt and the like. In addition to water, various liquid ingredients such as corn (maize) or malt syrup can be added. The farinaceous material may include cereal grains, cut grains, grits or flours from wheat, rice, corn, oats, barley, rye, or other cereal grains and mixtures thereof from that a transformed plant of the invention. The dough may then be processed into a desired shape through a process such as extrusion or stamping and further cooked using means such as a James cooker, an oven or an electrical discharge device.

Further provided is a method to sweeten a starch containing product without adding sweetener. The method comprises treating starch comprising at least one starch processing enzyme conditions to activate the at least one enzyme thereby digesting the starch to form a sugar thereby forming a treated (sweetened) starch, e.g., relative to the product produced by treating starch which does not comprise the hyperthermophilic enzyme. The starch of the invention is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one processing enzyme. Enzymes include  $\alpha$ -amylase,  $\alpha$ -glucosidase, glucoamylase, pullulanase, glucose isomerase, or any combination thereof. The enzyme may be hyperthermophilic and activated with heat. Preferred transformed plants include corn, soybean, rye, oat, barley, wheat, rice and sugar cane. The treated starch is then added to a product to produce a sweetened starch containing product, e.g., a farinaceous food product. Also provided is a sweetened starch containing product produced by the method.

The invention further provides a method to sweeten a polysaccharide containing fruit or vegetable comprising: treating a fruit or vegetable comprising at least one polysaccharide processing enzyme under conditions which activate the at least one enzyme, thereby processing the polysaccharide in the fruit or vegetable to form sugar, yielding a sweetened fruit or vegetable, e.g., relative to a fruit or vegetable from a plant which does not comprise the polysaccharide processing enzyme. The fruit or vegetable of the invention is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one polysaccharide processing enzyme.

Fruits and vegetables include potato, tomato, banana, squash, pea, and bean.

Enzymes include  $\alpha$ -amylase,  $\alpha$ -glucosidase, glucoamylase, pullulanase, glucose

isomerase, or any combination thereof. The enzyme may be hyperthermophilic.

i. Sweetening a polysaccharide containing plant or plant product

The method involves obtaining a plant that expresses a polysaccharide processing enzyme which converts a polysaccharide into a sugar as described above. Accordingly the enzyme is expressed in the plant and in the products of the plant, such as in a fruit or vegetable. In one embodiment, the enzyme is placed under the control of an inducible promoter such that expression of the enzyme may be induced by an external stimulus. Such inducible promoters and constructs are well known in the art and are described herein. Expression of the enzyme within the plant or product thereof causes polysaccharide contained within the plant or product thereof to be converted into sugar and to sweeten the plant or product thereof. In another embodiment, the polysaccharide processing enzyme is constitutively expressed. Thus, the plant or product thereof may be activated under conditions sufficient to activate the enzyme to convert the polysaccharides into sugar through the action of the enzyme to sweeten the plant or product thereof. As a result, this self-processing of the polysaccharide in the fruit or vegetable to form sugar yields a sweetened fruit or vegetable, e.g., relative to a fruit or vegetable from a plant which does not comprise the polysaccharide processing enzyme. The fruit or vegetable of the invention is obtained from a transformed plant, the genome of which is augmented with an expression cassette encoding the at least one polysaccharide processing enzyme. Fruits and vegetables include potato, tomato, banana, squash, pea, and bean. Enzymes include  $\alpha$ -amylase,  $\alpha$ -glucosidase, glucoamylase, pullulanase, glucose isomerase, or any combination thereof. The polysaccharide processing enzyme may be hyperthermophilic.

j. Isolation of starch from transformed grain that contains an enzyme which disrupts the endosperm matrix

The invention provides a method to isolate starch from a transformed grain wherein an enzyme is expressed that disrupts the endosperm matrix. The method involves obtaining a plant that expresses an enzyme which disrupts the endosperm matrix by modification of, for example, cell walls, non-starch polysaccharides and/or proteins. Examples of such enzymes include, but are not limited to, proteases, glucanases, thioredoxin, thioredoxin reductase, phytases, lipases, cellulases, beta glucosidases, xylanases and esterases. Such enzymes do not include any enzyme

that exhibits starch-degrading activity so as to maintain the integrity of the starch granules. The enzyme may be fused to a signal sequence that targets the enzyme to the starch granule. In one embodiment the grain is heat dried to activate the enzyme and inactivate the endogenous enzymes contained within the grain. The heat treatment causes activation of the enzyme, which acts to disrupt the endosperm matrix which is then easily separated from the starch granules. In another embodiment, the grain is steeped at low or high temperature, with high or low moisture content, with or without sulfur dioxide. The grain is then heat treated to disrupt the endosperm matrix and allow for easy separation of the starch granules. In another embodiment, proper temperature and moisture conditions are created to allow proteases to enter into the starch granules and degrade proteins contained within the granules. Such treatment would produce starch granules with high yield and little contaminating protein.

k. Syrup having a high sugar equivalent and use of the syrup to produce ethanol or a fermented beverage

The method involves obtaining a plant that expresses a polysaccharide processing enzyme which converts a polysaccharide into a sugar as described above. The plant, or product thereof, is steeped in an aqueous stream under conditions where the expressed enzyme converts polysaccharide contained within the plant, or product thereof, into dextrin, maltooligosaccharide, and/or sugar. The aqueous stream containing the dextrin, maltooligosaccharide, and/or sugar produced through conversion of the polysaccharide is then separated to produce a syrup having a high sugar equivalent. The method may or may not include an additional step of wet-milling the plant or product thereof to obtain starch granules. Examples of enzymes that may be used within the method include, but are not limited to,  $\alpha$ -amylase, glucoamylase, pullulanase and  $\alpha$ -glucosidase. The enzyme may be hyperthermophilic. Sugars produced according to the method include, but are not limited to, hexose, glucose and fructose. Examples of plants that may be used with the method include, but are not limited to, corn, wheat or barley. Examples of products of a plant that may be used include, but are not limited to, fruit, grain and vegetables. In one embodiment, the polysaccharide processing enzyme is placed under the control of an inducible promoter. Accordingly, prior to or during the steeping process, the promoter is induced to cause expression of the enzyme, which then provides for the conversion of

polysaccharide into sugar. Examples of inducible promoters and constructs containing them are well known in the art and are provided herein. Thus, where the polysaccharide processing is hyperthermophilic, the steeping is performed at a high temperature to activate the hyperthermophilic enzyme and inactivate endogenous enzymes found within the plant or product thereof. In another embodiment, a hyperthermophilic enzyme able to convert polysaccharide into sugar is constitutively expressed. This enzyme may or may not be targeted to a compartment within the plant through use of a signal sequence. The plant, or product thereof, is steeped under high temperature conditions to cause the conversion of polysaccharides contained within the plant into sugar.

Also provided is a method to produce ethanol or a fermented beverage from syrup having a high sugar equivalent. The method involves incubating the syrup with yeast under conditions that allow conversion of sugar contained within the syrup into ethanol or a fermented beverage. Examples of such fermented beverages include, but are not limited to, beer and wine. Fermentation conditions are well known in the art and are described in U.S. Patent No.: 4,929,452 and herein. Preferably the yeast is a high alcohol-tolerant and high-sugar tolerant strain of yeast such as *S. cerevisiae* ATCC No. 20867. The fermented product or fermented beverage may be distilled to isolate ethanol or a distilled beverage.

**I. Accumulation of hyperthermophilic enzyme in the cell wall of a plant**

The invention provides a method to accumulate a hyperthermophilic enzyme in the cell wall of a plant. The method involves expressing within a plant a hyperthermophilic enzyme that is fused to a cell wall targeting signal such that the targeted enzyme accumulates in the cell wall. Preferably the enzyme is able to convert polysaccharides into monosaccharides. Examples of targeting sequences include, but are not limited to, a cellulose or xylose binding domain. Examples of hyperthermophilic enzymes include those listed in SEQ ID NO: 1, 3, 5, 10, 13, 14, 15 or 16. Plant material containing cell walls may be added as a source of desired enzymes in a process to recover sugars from the feedstock or as a source of enzymes for the conversion of polysaccharides originating from other sources to monosaccharides. Additionally, the cell walls may serve as a source from which enzymes may be purified. Methods to purify enzymes are well known in the art and include, but are not limited to, gel filtration, ion-exchange chromatography, chromatofocusing, isoelectric focusing, affinity chromatography, FPLC,

HPLC, salt precipitation, dialysis, and the like. Accordingly, the invention also provides purified enzymes isolated from the cell walls of plants.

m. Method of preparing and isolating processing enzymes

In accordance with the present invention, recombinantly-produced processing enzymes of the present invention may be prepared by transforming plant tissue or plant cell comprising the processing enzyme of the present invention capable of being activated in the plant, selected for the transformed plant tissue or cell, growing the transformed plant tissue or cell into a transformed plant, and isolating the processing enzyme from the transformed plant or part thereof. The recombinantly-produced enzyme may be an  $\alpha$ -amylase, glucoamylase, glucose isomerase,  $\alpha$ -glucosidase, pullulanase, xylanase, protease, glucanase, beta glucosidase, esterase, lipase, or phytase. The enzyme may be encoded by the polynucleotide selected from any of SEQ ID NO: 2, 4, 6, 9, 19, 21, 25, 37, 39, 41, 43, 46, 48, 50, 52, 59, 61, 63, 65, 79, 81, 83, 85, 87, 89, 91, 93, 94, 95, 96, 97, or 99.

The invention will be further described by the following examples, which are not intended to limit the scope of the invention in any manner.

**Examples**

**Example 1**

**Construction of maize-optimized genes for hyperthermophilic starch-processing/isomerization enzymes**

The enzymes,  $\alpha$ -amylase, pullulanase,  $\alpha$ -glucosidase, and glucose isomerase, involved in starch degradation or glucose isomerization were selected for their desired activity profiles. These include, for example, minimal activity at ambient temperature, high temperature activity/stability, and activity at low pH. The corresponding genes were then designed by using maize preferred codons as described in U.S. Patent No. 5,625,136 and synthesized by Integrated DNA Technologies, Inc. (Coralville, IA).

The 797GL3  $\alpha$ -amylase, having the amino acid sequence SEQ ID NO:1, was selected for its hyperthermophilic activity. This enzyme's nucleic acid sequence was deduced and maize-

optimized as represented in SEQ ID NO:2. Similarly, the 6gp3 pullulanase was selected having the amino acid sequence set forth in SEQ ID NO:3. The nucleic acid sequence for the 6gp3 pullulanase was deduced and maize-optimized as represented in SEQ ID NO:4.

The amino acid sequence for malA  $\alpha$ -glucosidase from *Sulfolobus solfataricus* was obtained from the literature, J. Bact. 177:482-485 (1995); J. Bact. 180:1287-1295 (1998). Based on the published amino acid sequence of the protein (SEQ ID NO:5), the maize-optimized synthetic gene (SEQ ID NO:6) encoding the malA  $\alpha$ -glucosidase was designed.

Several glucose isomerase enzymes were selected. The amino acid sequence (SEQ ID NO:18) for glucose isomerase derived from *Thermotoga maritima* was predicted based on the published DNA sequence having Accession No. NC\_000853 and a maize-optimized synthetic gene was designed (SEQ ID NO: 19). Similarly the amino acid sequence (SEQ ID NO:20) for glucose isomerase derived from *Thermotoga neapolitana* was predicted based on the published DNA sequence from Appl. Envir. Microbiol. 61(5):1867-1875 (1995), Accession No. L38994. A maize-optimized synthetic gene encoding the *Thermotoga neapolitana* glucose isomerase was designed (SEQ ID NO:21).

## Example 2

### Expression of fusion of 797GL3 $\alpha$ -amylase and starch encapsulating region in *E. coli*

A construct encoding hyperthermophilic 797GL3  $\alpha$ -amylase fused to the starch encapsulating region (SER) from maize granule-bound starch synthase (waxy) was introduced and expressed in *E. coli*. The maize granule-bound starch synthase cDNA (SEQ ID NO:7) encoding the amino acid sequence (SEQ ID NO:8)(Klosgen RB, et al. 1986) was cloned as a source of a starch binding domain, or starch encapsulating region (SER). The full-length cDNA was amplified by RT-PCR from RNA prepared from maize seed using primers SV57 (5'AGCGAATTCATGGCGGCTCTGGCCACGT 3') (SEQ ID NO: 22) and SV58 (5'AGCTAAGCTTCAGGGCGCGGCCACGTTCT 3') (SEQ ID NO: 23) designed from GenBank Accession No. X03935. The complete cDNA was cloned into pBluescript as an EcoRI/HindIII fragment and the plasmid designated pNOV4022.

The C-terminal portion (encoded by bp 919-1818) of the waxy cDNA, including the starch-binding domain, was amplified from pNOV4022 and fused in-frame to the 3' end of the full-length maize-optimized 797GL3 gene (SEQ ID NO:2). The fused gene product, 797GL3/Waxy, having the nucleic acid SEQ ID NO:9 and encoding the amino acid sequence, SEQ ID NO:10, was cloned as an NcoI/XbaI fragment into pET28b (NOVAGEN, Madison, WI) that was cut with NcoI/NheI. The 797GL3 gene alone was also cloned into the pET28b vector as an NcoI/XbaI fragment.

The pET28/797GL3 and the pET28/797GL3/Waxy vectors were transformed into BL21/DE3 *E. coli* cells (NOVAGEN) and grown and induced according to the manufacturer's instruction. Analysis by PAGE/Coomassie staining revealed an induced protein in both extracts corresponding to the predicted sizes of the fused and unfused amylase, respectively.

Total cell extracts were analyzed for hyperthermophilic amylase activity as follows: 5 mg of starch was suspended in 20  $\mu$ l of water then diluted with 25  $\mu$ l of ethanol. The standard amylase positive control or the sample to be tested for amylase activity was added to the mixture and water was added to a final reaction volume of 500  $\mu$ l. The reaction was carried out at 80°C for 15-45 minutes. The reaction was then cooled down to room temperature, and 500  $\mu$ l of o-dianisidine and glucose oxidase/peroxidase mixture (Sigma) was added. The mixture was incubated at 37°C for 30 minutes. 500  $\mu$ l of 12 N sulfuric acid was added to stop the reaction. Absorbance at 540 nm was measured to quantitate the amount of glucose released by the amylase/sample. Assay of both the fused and unfused amylase extracts gave similar levels of hyperthermophilic amylase activity, whereas control extracts were negative. This indicated that the 797GL3 amylase was still active (at high temperatures) when fused to the C-terminal portion of the waxy protein.

### Example 3

#### Isolation of promoter fragments for endosperm-specific expression in maize.

The promoter and 5' noncoding region I (including the first intron) from the large subunit of *Zea mays* ADP-gpp (ADP-glucose pyrophosphorylase) was amplified as a 1515 base pair fragment (SEQ ID NO:11) from maize genomic DNA using primers designed from Genbank



accession M81603. The ADP-gpp promoter has been shown to be endosperm-specific (Shaw and Hannah, 1992).

The promoter from the *Zea mays*  $\gamma$ -zein gene was amplified as a 673 bp fragment (SEQ ID NO:12) from plasmid pGZ27.3 (obtained from Dr. Brian Larkins). The  $\gamma$ -zein promoter has been shown to be endosperm-specific (Torrent et al. 1997).

#### Example 4

##### Construction of transformation vectors for the 797GL3 hyperthermophilic $\alpha$ -amylase

Expression cassettes were constructed to express the 797GL3 hyperthermophilic amylase in maize endosperm with various targeting signals as follows:

pNOV6200 (SEQ ID NO:13) comprises the maize  $\gamma$ -zein N-terminal signal sequence (MRVLLVALALLALAASATS)(SEQ ID NO:17) fused to the synthetic 797GL3 amylase as described above in Example 1 for targeting to the endoplasmic reticulum and secretion into the apoplast (Torrent et al. 1997). The fusion was cloned behind the maize ADP-gpp promoter for expression specifically in the endosperm.

pNOV6201 (SEQ ID NO:14) comprises the  $\gamma$ -zein N-terminal signal sequence fused to the synthetic 797GL3 amylase with a C-terminal addition of the sequence SEKDEL for targeting to and retention in the endoplasmic reticulum (ER) (Munro and Pelham, 1987). The fusion was cloned behind the maize ADP-gpp promoter for expression specifically in the endosperm.

pNOV7013 comprises the  $\gamma$ -zein N-terminal signal sequence fused to the synthetic 797GL3 amylase with a C-terminal addition of the sequence SEKDEL for targeting to and retention in the endoplasmic reticulum (ER). PNOV7013 is the same as pNOV6201, except that the the maize  $\gamma$ -zein promoter (SEQ ID NO:12) was used instead of the maize ADP-spp promoter in order to express the fusion in the endosperm.

pNOV4029 (SEQ ID NO:15) comprises the waxy amyloplast targeting peptide (Klosgen et al., 1986) fused to the synthetic 797GL3 amylase for targeting to the amyloplast. The fusion was cloned behind the maize ADP-gpp promoter for expression specifically in the endosperm.

pNOV4031 (SEQ ID NO:16) comprises the waxy amyloplast targeting peptide fused to the synthetic 797GL3/waxy fusion protein for targeting to starch granules. The fusion was cloned behind the maize ADP-gpp promoter for expression specifically in the endosperm.

Additional constructs were made with these fusions cloned behind the maize  $\gamma$ -zein promoter to obtain higher levels of enzyme expression. All expression cassettes were moved into a binary vector for transformation into maize via *Agrobacterium* infection. The binary vector contained the phosphomannose isomerase (PMI) gene which allows for selection of transgenic cells with mannose. Transformed maize plants were either self-pollinated or outcrossed and seed was collected for analysis.

Additional constructs were made with the targeting signals described above fused to either 6gp3 pullulanase or to 340g12  $\alpha$ -glucosidase in precisely the same manner as described for the  $\alpha$ -amylase. These fusions were cloned behind the maize ADP-gpp promoter and/or the  $\gamma$ -zein promoter and transformed into maize as described above. Transformed maize plants were either self-pollinated or outcrossed and seed was collected for analysis.

Combinations of the enzymes can be produced either by crossing plants expressing the individual enzymes or by cloning several expression cassettes into the same binary vector to enable cotransformation.

### Example 5

#### Construction of plant transformation vectors for the 6GP3 thermophillic pullulanase

An expression cassette was constructed to express the 6GP3 thermophillic pullulanase in the endoplasmic reticulum of maize endosperm as follows:

pNOV7005 (SEQ ID NOs:24 and 25) comprises the maize  $\gamma$ -zein N-terminal signal sequence fused to the synthetic 6GP3 pullulanase with a C-terminal addition of the sequence SEKDEL for targeting to and retention in the ER. The amino acid peptide SEKDEL was fused to the C-terminal end of the enzymes using PCR with primers designed to amplify the synthetic gene and simultaneously add the 6 amino acids at the C-terminal end of the protein. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

### Example 6

#### Construction of plant transformation vectors for the malA

hyperthermophilic  $\alpha$ -glucosidase

Expression cassettes were constructed to express the *Sulfolobus solfataricus* malA hyperthermophilic  $\alpha$ -glucosidase in maize endosperm with various targeting signals as follows:

pNOV4831 (SEQ ID NO:26) comprises the maize  $\gamma$ -zein N-terminal signal sequence (MRVLLVALALLALAASATS)(SEQ ID NO:17) fused to the synthetic malA  $\alpha$ -glucosidase with a C-terminal addition of the sequence SEKDEL for targeting to and retention in the endoplasmic reticulum (ER) (Munro and Pelham, 1987). The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

pNOV4839 (SEQ ID NO:27) comprises the maize  $\gamma$ -zein N-terminal signal sequence (MRVLLVALALLALAASATS)(SEQ ID NO:17) fused to the synthetic malA  $\alpha$ -glucosidase for targeting to the endoplasmic reticulum and secretion into the apoplast (Torrent et al. 1997). The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

pNOV4837 comprises the maize  $\gamma$ -zein N-terminal signal sequence (MRVLLVALALLALAASATS)(SEQ ID NO:17) fused to the synthetic malA  $\alpha$ -glucosidase with a C-terminal addition of the sequence SEKDEL for targeting to and retention in the ER. The fusion was cloned behind the maize ADPgpp promoter for expression specifically in the endosperm. The amino acid sequence for this clone is identical to that of pNOV4831 (SEQ ID NO:26).

**Example 7****Construction of plant transformation vectors for the hyperthermophilic  
*Thermotoga maritima* and *Thermotoga neapolitana* glucose isomerases**

Expression cassettes were constructed to express the *Thermotoga maritima* and *Thermotoga neapolitana* hyperthermophilic glucose isomerases in maize endosperm with various targeting signals as follows:

pNOV4832 (SEQ ID NO:28) comprises the maize  $\gamma$ -zein N-terminal signal sequence (MRVLLVALALLALAASATS)(SEQ ID NO:17) fused to the synthetic *Thermotoga maritima* glucose isomerase with a C-terminal addition of the sequence SEKDEL for targeting to and retention in the ER. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

pNOV4833 (SEQ ID NO:29) comprises the maize  $\gamma$ -zein N-terminal signal sequence (MRVLLVALALLALAASATS)(SEQ ID NO:17) fused to the synthetic *Thermotoga neapolitana* glucose isomerase with a C-terminal addition of the sequence SEKDEL for targeting to and retention in the ER. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

pNOV4840 (SEQ ID NO:30) comprises the maize  $\gamma$ -zein N-terminal signal sequence (MRVLLVALALLALAASATS)(SEQ ID NO:17) fused to the synthetic *Thermotoga neapolitana* glucose isomerase for targeting to the endoplasmic reticulum and secretion into the apoplast (Torrent et al. 1997). The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

pNOV4838 comprises the maize  $\gamma$ -zein N-terminal signal sequence (MRVLLVALALLALAASATS)(SEQ ID NO:17) fused to the synthetic *Thermotoga neapolitana* glucose isomerase with a C-terminal addition of the sequence SEKDEL for targeting to and retention in the ER. The fusion was cloned behind the maize ADPgpp promoter for expression specifically in the endosperm. The amino acid sequence for this clone is identical to that of pNOV4833 (SEQ ID NO:29).

### Example 8

#### Construction of plant transformation vectors for the expression of the hyperthermophilic glucanase EglA

pNOV4800 (SEQ ID NO:58) comprises the barley alpha amylase AMY32b signal sequence (MGKNGNLCCFSLLLLLLAGLASGHQ)(SEQ ID NO:31) fused with the EglA mature protein sequence for localization to the apoplast. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

### Example 9

#### Construction of plant transformation vectors for the expression of multiple hyperthermophilic enzymes

pNOV4841 comprises a double gene construct of a 797GL3  $\alpha$ -amylase fusion and a 6GP3 pullulanase fusion. Both 797GL3 fusion (SEQ ID NO:33) and 6GP3 fusion (SEQ ID NO:34) possessed the maize  $\gamma$ -zein N-terminal signal sequence and SEKDEL sequence for targeting to and retention in the ER. Each fusion was cloned behind a separate maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

pNOV4842 comprises a double gene construct of a 797GL3  $\alpha$ -amylase fusion and a malA  $\alpha$ -glucosidase fusion. Both the 797GL3 fusion polypeptide (SEQ ID NO:35) and malA  $\alpha$ -glucosidase fusion polypeptide (SEQ ID NO:36) possess the maize  $\gamma$ -zein N-terminal signal sequence and SEKDEL sequence for targeting to and retention in the ER. Each fusion was cloned behind a separate maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

pNOV4843 comprises a double gene construct of a 797GL3  $\alpha$ -amylase fusion and a malA  $\alpha$ -glucosidase fusion. Both the 797GL3 fusion and malA  $\alpha$ -glucosidase fusion possess the maize  $\gamma$ -zein N-terminal signal sequence and SEKDEL sequence for targeting to and retention in the ER. The 797GL3 fusion was cloned behind the maize  $\gamma$ -zein promoter and the malA fusion was cloned behind the maize ADPgpp promoter for expression specifically in the endosperm. The amino acid sequences of the 797GL3 fusion and the malA fusion are identical to those of pNOV4842 (SEQ ID Nos: 35 and 36, respectively).

pNOV4844 comprises a triple gene construct of a 797GL3  $\alpha$ -amylase fusion, a 6GP3 pullulanase fusion, and a malA  $\alpha$ -glucosidase fusion. 797GL3, malA, and 6GP3 all possess the

maize  $\gamma$ -zein N-terminal signal sequence and SEKDEL sequence for targeting to and retention in the ER. The 797GL3 and malA fusions were cloned behind 2 separate maize  $\gamma$ -zein promoters, and the 6GP3 fusion was cloned behind the maize ADPgpp promoter for expression specifically in the endosperm. The amino acid sequences for the 797GL3 and malA fusions are identical to those of pNOV4842 (SEQ ID Nos: 35 and 36, respectively). The amino acid sequence for the 6GP3 fusion is identical to that of the 6GP3 fusion in pNOV4841 (SEQ ID NO:34).

All expression cassettes set forth in this Example as well as in the Examples that follow were moved into the binary vector pNOV2117 for transformation into maize via *Agrobacterium* infection. pNOV2117 contains the phosphomannose isomerase (PMI) gene allowing for selection of transgenic cells with mannose. pNOV2117 is a binary vector with both the pVS1 and ColE1 origins of replication. This vector contains the constitutive VirG gene from pAD1289 (Hansen, G., et al., PNAS USA 91:7603-7607 (1994), incorporated by reference herein) and a spectinomycin resistance gene from Tn7. Cloned into the polylinker between the right and left borders are the maize ubiquitin promoter, PMI coding region and nopaline synthase terminator of pNOV117 (Negrotto, D., et al., Plant Cell Reports 19:798-803 (2000), incorporated by reference herein). Transformed maize plants will either be self-pollinated or outcrossed and seed collected for analysis. Combinations of the different enzymes can be produced either by crossing plants expressing the individual enzymes or by transforming a plant with one of the multi-gene cassettes.

### Example 10

#### Construction of bacterial and *Pichia* expression vectors

Expression cassettes were constructed to express the hyperthermophilic  $\alpha$ -glucosidase and glucose isomerases in either *Pichia* or bacteria as follows:

pNOV4829 (SEQ ID NOS: 37 and 38) comprises a synthetic *Thermotoga maritima* glucose isomerase fusion with ER retention signal in the bacterial expression vector pET29a. The glucose isomerase fusion gene was cloned into the NcoI and SacI sites of pET29a, which results in the addition of an N-terminal S-tag for protein purification.

pNOV4830 (SEQ ID NOS: 39 and 40) comprises a synthetic *Thermotoga neapolitana* glucose isomerase fusion with ER retention signal in the bacterial expression vector pET29a.

The glucose isomerase fusion gene was cloned into the NcoI and SacI sites of pET29a, which results in the addition of an N-terminal S-tag for protein purification.

pNOV4835 (SEQ ID NO: 41 and 42) comprises the synthetic *Thermotoga maritima* glucose isomerase gene cloned into the BamHI and EcoRI sites of the bacterial expression vector pET28C. This resulted in the fusion of a His-tag (for protein purification) to the N-terminal end of the glucose isomerase.

pNOV4836 (SEQ ID NO: 43 AND 44) comprises the synthetic *Thermotoga neapolitana* glucose isomerase gene cloned into the BamHI and EcoRI sites of the bacterial expression vector pET28C. This resulted in the fusion of a His-tag (for protein purification) to the N-terminal end of the glucose isomerase.

### Example 11

Transformation of immature maize embryos was performed essentially as described in Negrotto et al., Plant Cell Reports 19: 798-803. For this example, all media constituents are as described in Negrotto et al., *supra*. However, various media constituents described in the literature may be substituted.

#### A. Transformation plasmids and selectable marker

The genes used for transformation were cloned into a vector suitable for maize transformation. Vectors used in this example contained the phosphomannose isomerase (PMI) gene for selection of transgenic lines (Negrotto et al. (2000) Plant Cell Reports 19: 798-803).

#### B. Preparation of *Agrobacterium tumefaciens*

*Agrobacterium* strain LBA4404 (pSB1) containing the plant transformation plasmid was grown on YEP (yeast extract (5 g/L), peptone (10g/L), NaCl (5g/L), 15g/l agar, pH 6.8) solid medium for 2 – 4 days at 28°C. Approximately  $0.8 \times 10^9$  *Agrobacterium* were suspended in LS-inf media supplemented with 100  $\mu$ M As (Negrotto *et al.*, (2000) Plant Cell Rep 19: 798-803). Bacteria were pre-induced in this medium for 30-60 minutes.

### C. Inoculation

Immature embryos from A188 or other suitable genotype were excised from 8 – 12 day old ears into liquid LS-inf + 100  $\mu$ M As. Embryos were rinsed once with fresh infection medium. *Agrobacterium* solution was then added and embryos were vortexed for 30 seconds and allowed to settle with the bacteria for 5 minutes. The embryos were then transferred scutellum side up to LSAs medium and cultured in the dark for two to three days. Subsequently, between 20 and 25 embryos per petri plate were transferred to LSDc medium supplemented with cefotaxime (250 mg/l) and silver nitrate (1.6 mg/l) and cultured in the dark for 28°C for 10 days.

### D. Selection of transformed cells and regeneration of transformed plants

Immature embryos producing embryogenic callus were transferred to LSD1M0.5S medium. The cultures were selected on this medium for 6 weeks with a subculture step at 3 weeks. Surviving calli were transferred to Reg1 medium supplemented with mannose. Following culturing in the light (16 hour light/ 8 hour dark regiment), green tissues were then transferred to Reg2 medium without growth regulators and incubated for 1-2 weeks. Plantlets are transferred to Magenta GA-7 boxes (Magenta Corp, Chicago Ill.) containing Reg3 medium and grown in the light. After 2-3 weeks, plants were tested for the presence of the PMI genes and other genes of interest by PCR. Positive plants from the PCR assay were transferred to the greenhouse.

### Example 12

#### Analysis of T1 seed from maize plants expressing the $\alpha$ -amylase targeted to apoplast or to the ER

T1 seed from self-pollinated maize plants transformed with either pNOV6200 or pNOV6201 as described in Example 4 were obtained. Starch accumulation in these kernels appeared to be normal, based on visual inspection and on normal staining for starch with an iodine solution prior to any exposure to high temperature. Immature kernels were dissected and purified endosperms were placed individually in microfuge tubes and immersed in 200  $\mu$ l of 50 mM NaPO<sub>4</sub> buffer. The tubes were placed in an 85°C water bath for 20 minutes, then cooled on ice. Twenty microliters of a 1% iodine solution was added to each tube and mixed. Approximately 25% of the segregating kernels stained normally for starch. The remaining 75% failed to stain, indicating that the starch had been degraded into low molecular weight sugars that



do not stain with iodine. It was found that the T1 kernels of pNOV6200 and pNOV6201 were self-hydrolyzing the corn starch. There was no detectable reduction in starch following incubation at 37°C.

Expression of the amylase was further analyzed by isolation of the hyperthermophilic protein fraction from the endosperm followed by PAGE/Coomassie staining. A segregating protein band of the appropriate molecular weight (50 kD) was observed. These samples are subjected to an  $\alpha$ -amylase assay using commercially available dyed amylose (AMYLZYME, from Megazyme, Ireland). High levels of hyperthermophilic amylase activity correlated with the presence of the 50 kD protein.

It was further found that starch in kernels from a majority of transgenic maize, which express hyperthermophilic  $\alpha$ -amylase, targeted to the amyloplast, is sufficiently active at ambient temperature to hydrolyze most of the starch if the enzyme is allowed to be in direct contact with a starch granule. Of the eighty lines having hyperthermophilic  $\alpha$ -amylase targeted to the amyloplast, four lines were identified that accumulate starch in the kernels. Three of these lines were analyzed for thermostable  $\alpha$ -amylase activity using a colorimetric amylazyme assay (Megazyme). The amylase enzyme assay indicated that these three lines had low levels of thermostable amylase activity. When purified starch from these three lines was treated with appropriate conditions of moisture and heat, the starch was hydrolyzed indicating the presence of adequate levels of  $\alpha$ -amylase to facilitate the auto-hydrolysis of the starch prepared from these lines.

T1 seed from multiple independent lines of both pNOV6200 and pNOV6201 transformants was obtained. Individual kernels from each line were dissected and purified endosperms were homogenized individually in 300  $\mu$ l of 50 mM NaPO<sub>4</sub> buffer. Aliquots of the endosperm suspensions were analyzed for  $\alpha$ -amylase activity at 85°C. Approximately 80% of the lines segregate for hyperthermophilic activity (See Figures 1A, 1B, and 2).

Kernels from wild type plants or plants transformed with pNOV6201 were heated at 100°C for 1, 2, 3, or 6 hours and then stained for starch with an iodine solution. Little or no starch was detected in mature kernels after 3 or 6 hours, respectively. Thus, starch in mature

kernels from transgenic maize which express hyperthermophilic amylase that is targeted to the endoplasmic reticulum was hydrolyzed when incubated at high temperature.

In another experiment, partially purified starch from mature T1 kernels from pNOV6201 plants that were steeped at 50°C for 16 hours was hydrolyzed after heating at 85°C for 5 minutes. This illustrated that the  $\alpha$ -amylase targeted to the endoplasmic reticulum binds to starch after grinding of the kernel, and is able to hydrolyze the starch upon heating. Iodine staining indicated that the starch remains intact in mature seeds after the 16 hour steep at 50°C.

In another experiment, segregating, mature kernels from plants transformed with pNOV6201 were heated at 95°C for 16 hours and then dried. In seeds expressing the hyperthermophilic  $\alpha$ -amylase, the hydrolysis of starch to sugar resulted in a wrinkled appearance following drying.

### **Example 13**

#### **Analysis of T1 seed from maize plants expressing the $\alpha$ -amylase targeted to the amyloplast**

T1 seed from self-pollinated maize plants transformed with either pNOV4029 or pNOV4031 as described in Example 4 was obtained. Starch accumulation in kernels from these lines was clearly not normal. All lines segregated, with some variation in severity, for a very low or no starch phenotype. Endosperm purified from immature kernels stained only weakly with iodine prior to exposure to high temperatures. After 20 minutes at 85°C, there was no staining. When the ears were dried, the kernels shriveled up. This particular amylase clearly had sufficient activity at greenhouse temperatures to hydrolyze starch if allowed to be in direct contact with the granule

### **Example 14**

#### **Fermentation of grain from maize plants expressing $\alpha$ -amylase**

100% Transgenic grain 85°C vs. 95°C, varied liquefaction time.

Transgenic corn (pNOV6201) that contains a thermostable  $\alpha$ -amylase performs well in fermentation without addition of exogenous  $\alpha$ -amylase, requires much less time for liquefaction and results in more complete solubilization of starch. Laboratory scale fermentations were

performed by a protocol with the following steps (detailed below): 1) grinding, 2) moisture analysis, 3) preparation of a slurry containing ground corn, water, backset and  $\alpha$ -amylase, 4) liquefaction and 5) simultaneous saccharification and fermentation (SSF). In this example the temperature and time of the liquefaction step were varied as described below. In addition the transgenic corn was liquefied with and without exogenous  $\alpha$ -amylase and the performance in ethanol production compared to control corn treated with commercially available  $\alpha$ -amylase.

The transgenic corn used in this example was made in accordance with the procedures set out in Example 4 using a vector comprising the  $\alpha$ -amylase gene and the PMI selectable marker, namely pNOV6201. The transgenic corn was produced by pollinating a commercial hybrid (N3030BT) with pollen from a transgenic line expressing a high level of thermostable  $\alpha$ -amylase. The corn was dried to 11% moisture and stored at room temperature. The  $\alpha$ -amylase content of the transgenic corn flour was 95 units/g where 1 unit of enzyme generates 1 micromole reducing ends per min from corn flour at 85 °C in pH 6.0 MES buffer. The control corn that was used was a yellow dent corn known to perform well in ethanol production.

1) Grinding: Transgenic corn (1180 g) was ground in a Perten 3100 hammer mill equipped with a 2.0 mm screen thus generating transgenic corn flour. Control corn was ground in the same mill after thoroughly cleaning to prevent contamination by the transgenic corn.

2) Moisture analysis: Samples (20 g) of transgenic and control corn were weighed into aluminum weigh boats and heated at 100 °C for 4 h. The samples were weighed again and the moisture content calculated from the weight loss. The moisture content of transgenic flour was 9.26%; that of the control flour was 12.54%.

3) Preparation of slurries: The composition of slurries was designed to yield a mash with 36% solids at the beginning of SSF. Control samples were prepared in 100 ml plastic bottles and contained 21.50 g of control corn flour, 23 ml of de-ionized water, 6.0 ml of backset (8% solids by weight), and 0.30 ml of a commercially available  $\alpha$ -amylase diluted 1/50 with water. The  $\alpha$ -amylase dose was chosen as representative of industrial usage. When assayed under the conditions described above for assay of the transgenic  $\alpha$ -amylase, the control  $\alpha$ -amylase dose was 2 U/g corn flour. pH was adjusted to 6.0 by addition of ammonium hydroxide. Transgenic samples were prepared in the same fashion but contained 20 g of corn flour because of the lower

moisture content of transgenic flour. Slurries of transgenic flour were prepared either with  $\alpha$ -amylase at the same dose as the control samples or without exogenous  $\alpha$ -amylase.

4) Liquefaction: The bottles containing slurries of transgenic corn flour were immersed in water baths at either 85 °C or 95 °C for times of 5, 15, 30, 45 or 60 min. Control slurries were incubated for 60 min at 85 °C. During the high temperature incubation the slurries were mixed vigorously by hand every 5 min. After the high temperature step the slurries were cooled on ice.

5) Simultaneous saccharification and fermentation: The mash produced by liquefaction was mixed with glucoamylase (0.65 ml of a 1/50 dilution of a commercially available L-400 glucoamylase), protease (0.60 ml of a 1,000-fold dilution of a commercially available protease), 0.2 mg Lactocide & urea (0.85 ml of a 10-fold dilution of 50% Urea Liquor). A hole was cut into the cap of the 100 ml bottle containing the mash to allow CO<sub>2</sub> to vent. The mash was then inoculated with yeast (1.44 ml) and incubated in a water bath set at 90 F. After 24 hours of fermentation the temperature was lowered to 86 F; at 48 hours it was set to 82 F.

Yeast for inoculation was propagated by preparing a mixture that contained yeast (0.12 g) with 70 grams maltodextrin, 230 ml water, 100 ml backset, glucoamylase (0.88 ml of a 10-fold dilution of a commercially available glucoamylase), protease (1.76 ml of a 100-fold dilution of a commercially available enzyme), urea (1.07 grams), penicillin (0.67 mg) and zinc sulfate (0.13 g). The propagation culture was initiated the day before it was needed and was incubated with mixing at 90°F.

At 24, 48 & 72 hour samples were taken from each fermentation vessel, filtered through 0.2  $\mu$ m filters and analyzed by HPLC for ethanol & sugars. At 72 h samples were analyzed for total dissolved solids and for residual starch.

HPLC analysis was performed on a binary gradient system equipped with refractive index detector, column heater & Bio-Rad Aminex HPX-87H column. The system was equilibrated with 0.005 M H<sub>2</sub>SO<sub>4</sub> in water at 1 ml/min. Column temperature was 50 °C. Sample injection volume was 5  $\mu$ l; elution was in the same solvent. The RI response was calibrated by injection of known standards. Ethanol and glucose were both measured in each injection.

Residual starch was measured as follows. Samples and standards were dried at 50°C in an oven, then ground to a powder in a sample mill. The powder (0.2 g) was weighed into a 15

ml graduated centrifuge tube. The powder was washed 3 times with 10 ml aqueous ethanol (80% v/v) by vortexing followed by centrifugation and discarding of the supernatant. DMSO (2.0 ml) was added to the pellet followed by 3.0 ml of a thermostable alpha-amylase (300 units) in MOPS buffer. After vigorous mixing, the tubes were incubated in a water bath at 85°C for 60 min. During the incubation, the tubes were mixed four times. The samples were cooled and 4.0 ml sodium acetate buffer (200 mM, pH 4.5) was added followed by 0.1 ml of glucoamylase (20 U). Samples were incubated at 50°C for 2 hours, mixed, then centrifuged for 5 min at 3,500 rpm. The supernatant was filtered through a 0.2 um filter and analyzed for glucose by the HPLC method described above. An injection size of 50 µl was used for samples with low residual starch (<20% of solids).

Results Transgenic corn performed well in fermentation without added  $\alpha$ -amylase. The yield of ethanol at 72 hours was essentially the same with or without exogenous  $\alpha$ -amylase as shown in Table I. These data also show that a higher yield of ethanol is achieved when the liquefaction temperature is higher; the present enzyme expressed in the transgenic corn has activity at higher temperatures than other enzymes used commercially such as the *Bacillus liquefaciens*  $\alpha$ -amylase.

Table I

Liquefaction temp °C	Liquefaction time min.	Exogenous $\alpha$ -amylase	# replicates	Mean Ethanol % v/v	Std. Dev. % v/v
85	60	Yes	4	17.53	0.18
85	60	No	4	17.78	0.27
95	60	Yes	2	18.22	ND
95	60	No	2	18.25	ND

When the liquefaction time was varied, it was found that the liquefaction time required for efficient ethanol production was much less than the hour required by the conventional process. Figure 3 shows that the ethanol yield at 72 hours fermentation was almost unchanged from 15 min to 60 min liquefaction. In addition liquefaction at 95°C gave more ethanol at each time point than at the 85°C liquefaction. This observation demonstrates the process improvement achieved by use of a hyperthermophilic enzyme.

The control corn gave a higher final ethanol yield than the transgenic corn, but the control was chosen because it performs very well in fermentation. In contrast the transgenic corn has a genetic background chosen to facilitate transformation. Introducing the  $\alpha$ -amylase-trait into elite corn germplasm by well-known breeding techniques should eliminate this difference.

Examination of the residual starch levels of the beer produced at 72 hours (Figure 4) shows that the transgenic  $\alpha$ -amylase results in significant improvement in making starch available for fermentation; much less starch was left over after fermentation.

Using both ethanol levels and residual starch levels the optimal liquefaction times were 15 min at 95°C and 30 min at 85°C. In the present experiments these times were the total time that the fermentation vessels were in the water bath and thus include a time period during which the temperature of the samples was increasing from room temperature to 85°C or 95°C. Shorter liquefaction times may be optimal in large scale industrial processes that rapidly heat the mash by use of equipment such as jet cookers. Conventional industrial liquefaction processes require holding tanks to allow the mash to be incubated at high temperature for one or more hours. The

present invention eliminates the need for such holding tanks and will increase the productivity of liquefaction equipment.

One important function of  $\alpha$ -amylase in fermentation processes is to reduce the viscosity of the mash. At all time points the samples containing transgenic corn flour were markedly less viscous than the control sample. In addition the transgenic samples did not appear to go through the gelatinous phase observed with all control samples; gelatinization normally occurs when corn slurries are cooked. Thus having the  $\alpha$ -amylase distributed throughout the fragments of the endosperm gives advantageous physical properties to the mash during cooking by preventing formation of large gels that slow diffusion and increase the energy costs of mixing and pumping the mash.

The high dose of  $\alpha$ -amylase in the transgenic corn may also contribute to the favorable properties of the transgenic mash. At 85°C, the  $\alpha$ -amylase activity of the transgenic corn was many times greater activity than the of the dose of exogenous  $\alpha$ -amylase used in controls. The latter was chosen as representative of commercial use rates.

#### **Example 15**

##### Effective function of transgenic corn when mixed with control corn

Transgenic corn flour was mixed with control corn flour in various levels from 5% to 100% transgenic corn flour. These were treated as described in Example 14. The mashes containing transgenically expressed  $\alpha$ -amylase were liquefied at 85 °C for 30 min or at 95 °C for 15 min; control mashes were prepared as described in Example 14 and were liquefied at 85 °C for 30 or 60 min (one each) or at 95 °C for 15 or 60 min (one each).

The data for ethanol at 48 and 72 hours and for residual starch are given in Table 2. The ethanol levels at 48 hours are graphed in Figure 5; the residual starch determinations are shown in Figure 6. These data show that transgenically expressed thermostable  $\alpha$ -amylase gives very good performance in ethanol production even when the transgenic grain is only a small portion (as low as 5%) of the total grain in the mash. The data also show that residual starch is markedly lower than in control mash when the transgenic grain comprises at least 40% of the total grain.

Table 2

Transgenic grain wt %	85 °C Liquefaction			95 °C Liquefaction		
	Residual Starch	Ethanol	Ethanol	Residual Starch	Ethanol	Ethanol
		48 h	% v/v 72 h		48 h	% v/v 72 h
100	3.58	16.71	18.32	4.19	17.72	21.14
80	4.06	17.04	19.2	3.15	17.42	19.45
60	3.86	17.16	19.67	4.81	17.58	19.57
40	5.14	17.28	19.83	8.69	17.56	19.51
20	8.77	17.11	19.5	11.05	17.71	19.36
10	10.03	18.05	19.76	10.8	17.83	19.28
5	10.67	18.08	19.41	12.44	17.61	19.38
0*	7.79	17.64	20.11	11.23	17.88	19.87

\* Control samples . Values the average of 2 determinations

#### Example 16

##### Ethanol production as a function of liquefaction pH using

##### transgenic corn at a rate of 1.5 to 12 % of total corn

Because the transgenic corn performed well at a level of 5-10% of total corn in a fermentation, an additional series of fermentations in which the transgenic corn comprised 1.5 to 12% of the total corn was performed. The pH was varied from 6.4 to 5.2 and the  $\alpha$ -amylase enzyme expressed in the transgenic corn was optimized for activity at lower pH than is conventionally used industrially.

The experiments were performed as described in Example 15 with the following exceptions:

- 1). Transgenic flour was mixed with control flour as a percent of total dry weight at the levels ranging from 1.5% to 12.0%.
- 2). Control corn was N3030BT which is more similar to the transgenic corn than the control used in examples 14 and 15.
- 3). No exogenous  $\alpha$ -amylase was added to samples containing transgenic flour.



- 4). Samples were adjusted to pH 5.2, 5.6, 6.0 or 6.4 prior to liquefaction. At least 5 samples spanning the range from 0% transgenic corn flour to 12% transgenic corn flour were prepared for each pH.
- 5). Liquefaction for all samples was performed at 85 °C for 60 min.

The change in ethanol content as a function of fermentation time are shown in Figure 7. This figure shows the data obtained from samples that contained 3% transgenic corn. At the lower pH, the fermentation proceeds more quickly than at pH 6.0 and above; similar behavior was observed in samples with other doses of transgenic grain. The pH profile of activity of the transgenic enzyme combined with the high levels of expression will allow lower pH liquefactions resulting in more rapid fermentations and thus higher throughput than is possible at the conventional pH 6.0 process.

The ethanol yields at 72 hours are shown in Figure 8. As can be seen, on the basis of ethanol yield, the results showed little dependence on the amount of transgenic grain included in the sample. Thus the grain contains abundant amylase to facilitate fermentative production of ethanol. It also demonstrates that lower pH of liquefaction results in higher ethanol yield.

The viscosity of the samples after liquefaction was monitored and it was observed that at pH 6.0, 6% transgenic grain is sufficient for adequate reduction in viscosity. At pH 5.2 and 5.6, viscosity is equivalent to that of the control at 12% transgenic grain, but not at lower percentages of transgenic grain.

### **Example 17**

#### **Production of fructose from corn flour using thermophilic enzymes**

Corn that expresses the hyperthermophilic  $\alpha$ -amylase, 797GL3, was shown to facilitate production of fructose when mixed with an  $\alpha$ -glucosidase (MaA) and a xylose isomerase (XylA).

Seed from pNOV6201 transgenic plants expressing 797GL3 were ground to a flour in a Kleco cell thus creating amylase flour. Non-transgenic corn kernels were ground in the same manner to generate control flour.

The  $\alpha$ -glucosidase, MaA (from *S. solfataricus*), was expressed in *E. coli*. Harvested bacteria were suspended in 50 mM potassium phosphate buffer pH 7.0 containing 1 mM 4-(2-

aminoethyl)benzenesulfonyl fluoride then lysed in a French pressure cell. The lysate was centrifuged at 23,000 x g for 15 min at 4°C. The supernatant solution was removed, heated to 70°C for 10 min, cooled on ice for 10 min, then centrifuged at 34,000 x g for 30 min at 4°C. The supernatant solution was removed and the MalA concentrated two-fold in centricon 10 devices. The filtrate of the centricon 10 step was retained for use as a negative control for MalA.

Xylose (glucose) isomerase was prepared by expressing the xylA gene of *T. neapolitana* in *E. coli*. Bacteria were suspended in 100 mM sodium phosphate pH 7.0 and lysed by passage through a French pressure cell. After precipitation of cell debris, the extract was heated at 80°C for 10 min then centrifuged. The supernatant solution contained the XylA enzymatic activity. An empty-vector control extract was prepared in parallel with the XylA extract.

Corn flour (60 mg per sample) was mixed with buffer and extracts from *E. coli*. As indicated in Table 3, samples contained amylase corn flour (amylase) or control corn flour (control), 50 µl of either MalA extract (+) or filtrate (-), and 20 µl of either XylA extract (+) or empty vector control (-). All samples also contained 230 µl of 50mM MOPS, 10mM MgSO<sub>4</sub>, and 1 mM CoCl<sub>2</sub>; pH of the buffer was 7.0 at room temperature.

Samples were incubated at 85°C for 18 hours. At the end of the incubation time, samples were diluted with 0.9 ml of 85°C water and centrifuged to remove insoluble material. The supernatant fraction was then filtered through a Centricon3 ultrafiltration device and analyzed by HPLC with ELSD detection.

The gradient HPLC system was equipped with Astec Polymer Amino Column, 5 micron particle size, 250 X 4.6 mm and an Alltech ELSD 2000 detector. The system was pre-equilibrated with a 15:85 mixture of water:acetonitrile. The flow rate was 1 ml/min. The initial conditions were maintained for 5 min after injection followed by a 20 min gradient to 50:50 water:acetonitrile followed by 10 minutes of the same solvent. The system was washed with 20 min of 80:20 water:acetonitrile and then re-equilibrated with the starting solvent. Fructose was eluted at 5.8 min and glucose at 8.7 min.

Table 3

Sample	Corn flour	MalA	XylA	fructose peak area x 10 <sup>-6</sup>	glucose peak area x 10 <sup>-6</sup>
1	amylase	+	+	25.9	110.3
2	amylase	-	+	7.0	12.4
3	amylase	+	-	0.1	147.5
4	amylase	-	-	0	25.9
5	control	+	+	0.8	0.5
6	control	-	+	0.3	0.2
7	control	+	-	1.3	1.7
8	control	-	-	0.2	0.3

The HPLC results also indicated the presence of larger maltooligosaccharides in all samples containing the  $\alpha$ -amylase. These results demonstrate that the three thermophilic enzymes can function together to produce fructose from corn flour at a high temperature.

#### Example 18

##### Amylase Flour with Isomerase

In another example, amylase flour was mixed with purified MalA and each of two bacterial xylose isomerases: XylA of *T. maritima*, and an enzyme designated BD8037 obtained from Diversa. Amylase flour was prepared as described in Example 18.

*S. solfataricus* MalA with a 6His purification tag was expressed in *E. coli*. Cell lysate was prepared as described in Example 18, then purified to apparent homogeneity using a nickel affinity resin (Probond, Invitrogen) and following the manufacturer's instructions for native protein purification.

*T. maritima* XylA with the addition of an S tag and an ER retention signal was expressed in *E. coli* and prepared in the same manner as the *T. neapolitana* XylA described in Example 18.

Xylose isomerase BD8037 was obtained as a lyophilized powder and resuspended in 0.4x the original volume of water.

Amylase corn flour was mixed with enzyme solutions plus water or buffer. All reactions contained 60 mg amylase flour and a total of 600 $\mu$ l of liquid. One set of reactions was buffered with 50 mM MOPS, pH 7.0 at room temperature, plus 10mM MgSO<sub>4</sub> and 1 mM CoCl<sub>2</sub>; in a second set of reactions the metal-containing buffer solution was replaced by water. Isomerase enzyme amounts were varied as indicated in Table 4. All reactions were incubated for 2 hours at 90°C. Reaction supernatant fractions were prepared by centrifugation. The pellets were washed with an additional 600 $\mu$ l H<sub>2</sub>O and recentrifuged. The supernatant fractions from each reaction were combined, filtered through a Centricon 10, and analyzed by HPLC with ELSD detection as described in Example 17. The amounts of glucose and fructose observed are graphed in Figure 15.

Table 4

Sample	Amylase flour	Mal A	Isomerase
1	60 mg	+	none
2	60 mg	+	<i>T. maritima</i> , 100 $\mu$ l
3	60 mg	+	<i>T. maritima</i> , 10 $\mu$ l
4	60 mg	+	<i>T. maritima</i> , 2 $\mu$ l
5	60 mg	+	BD8037, 100 $\mu$ l
7	60mg	+	BD8037, 2 $\mu$ l
C	60 mg	none	none

With each of the isomerases, fructose was produced from corn flour in a dose-dependent manner when  $\alpha$ -amylase and  $\alpha$ -glucosidase were present in the reaction. These results demonstrate that the grain-expressed amylase 797GL3 can function with MalA and a variety of different thermophilic isomerases, with or without added metal ions, to produce fructose from corn flour at a high temperature. In the presence of added divalent metal ions, the isomerases can achieve the predicted fructose: glucose equilibrium at 90°C of approximately 55% fructose.

This would be an improvement over the current process using mesophilic isomerases, which requires a chromatographic separation to increase the fructose concentration.

### Example 19

#### Expression of a pullulanase in corn

Transgenic plants that were homozygous for either pNOV7013 or pNOV7005 were crossed to generate transgenic corn seed expressing both the 797GL3  $\alpha$ -amylase and 6GP3 pullulanase.

T1 or T2 seed from self-pollinated maize plants transformed with either pNOV 7005 or pNOV 4093 were obtained. pNOV4093 is a fusion of the maize optimized synthetic gene for 6GP3 (SEQ ID: 3,4) with the amyloplast targeting sequence (SEQ ID NO: 7,8) for localization of the fusion protein to the amyloplast. This fusion protein is under the control of the ADPgpp promoter (SEQ ID NO:11) for expression specifically in the endosperm. The pNOV7005 construct targets the expression of the pullulanase in the endoplasmic reticulum of the endosperm. Localization of this enzyme in the ER allows normal accumulation of the starch in the kernels. Normal staining for starch with an iodine solution was also observed, prior to any exposure to high temperature.

As described in the case of  $\alpha$ -amylase the expression of pullulanase targeted to the amyloplast (pNOV4093) resulted in abnormal starch accumulation in the kernels. When the corn-ears are dried, the kernels shriveled up. Apparently, this thermophilic pullulanase is sufficiently active at low temperatures and hydrolyzes starch if allowed to be in direct contact with the starch granules in the seed endosperm.

Enzyme preparation or extraction of the enzyme from corn-flour: The pullulanase enzyme was extracted from the transgenic seeds by grinding them in Kleco grinder, followed by incubation of the flour in 50mM NaOAc pH 5.5 buffer for 1 hr at RT, with continuous shaking. The incubated mixture was then spun for 15min. at 14000 rpm. The supernatant was used as enzyme source.

Pullulanase assay: The assay reaction was carried out in 96-well plate. The enzyme extracted from the corn flour (100  $\mu$ l) was diluted 10 fold with 900  $\mu$ l of 50mM NaOAc pH5.5 buffer, containing 40 mM  $\text{CaCl}_2$ . The mixture was vortexed, 1 tablet of Limit-Dextrizyme

(azurine-crosslinked-pullulan, from Megazyme) was added to each reaction mixture and incubated at 75 °C for 30 min (or as mentioned). At the end of the incubation the reaction mixtures were spun at 3500 rpm for 15 min. The supernatants were diluted 5 fold and transferred into 96-well flat bottom plate for absorbance measurement at 590 nm. Hydrolysis of azurine-crosslinked-pullulan substrate by the pullulanase produces water-soluble dye fragments and the rate of release of these (measured as the increase in absorbance at 590 nm) is related directly to enzyme activity.

Figure 9 shows the analysis of T2 seeds from different events transformed with pNOV 7005. High expression of pullulanase activity, compared to the non-transgenic control, can be detected in a number of events.

To a measured amount (~100 µg) of dry corn flour from transgenic (expressing pullulanase, or amylase or both the enzymes) and / or control (non-transgenic) 1000 µl of 50 mM NaOAc pH 5.5 buffer containing 40 mM CaCl<sub>2</sub> was added. The reaction mixtures were vortexed and incubated on a shaker for 1 hr. The enzymatic reaction was started by transferring the incubation mixtures to high temperature (75 °C, the optimum reaction temperature for pullulanase or as mentioned in the figures) for a period of time as indicated in the figures. The reactions were stopped by cooling them down on ice. The reaction mixtures were then centrifuged for 10 min. at 14000 rpm. An aliquot (100 µl) of the supernatant was diluted three fold, filtered through 0.2-micron filter for HPLC analysis.

The samples were analyzed by HPLC using the following conditions:

Column: Alltech Prevail Carbohydrate ES 5 micron 250 X 4.6 mm

Detector: Alltech ELSD 2000

Pump: Gilson 322

Injector: Gilson 215 injector/diluter

Solvents: HPLC grade Acetonitrile (Fisher Scientific) and Water (purified by Waters Millipore System)

Gradient used for oligosaccharides of low degree of polymerization (DP 1-15).

Time	%Water	%Acetonitrile
0	15	85
5	15	85
25	50	50
35	50	50
36	80	20
55	80	20
56	15	85
76	15	85

Gradient used for saccharides of high degree of polymerization (DP 20 – 100 and above).

Time	%Water	%Acetonitrile
0	35	65
60	85	15
70	85	15
85	35	65
100	35	65

System used for data analysis: Gilson Unipoint Software System Version 3.2

Figures 10A and 10B show the HPLC analysis of the hydrolytic products generated by expressed pullulanase from starch in the transgenic corn flour. Incubation of the flour of pullulanase expressing corn in reaction buffer at 75 °C for 30 minutes results in production of medium chain oligosaccharides (DP ~10-30) and short amylose chains (DP ~ 100 –200) from cornstarch. This figure also shows the dependence of pullulanase activity on presence of calcium ions.

Transgenic corn expressing pullulanase can be used to produce modified-starch/dextrin that is debranched ( $\alpha$ 1-6 linkages cleaved) and hence will have high level of amylose/straight chain dextrin. Also depending on the kind of starch (*e.g.* waxy, high amylose *etc.*) used the

chain length distribution of the amylose/dextrin generated by the pullulanase will vary, and so will the property of the modified-starch/dextrin.

Hydrolysis of  $\alpha$  1-6 linkage was also demonstrated using pullulan as the substrate. The pullulanase isolated from corn flour efficiently hydrolyzed pullulan. HPLC analysis (as described) of the product generated at the end of incubation showed production of maltotriose, as expected, due to the hydrolysis of the  $\alpha$  1-6 linkages in the pullulan molecules by the enzyme from the corn.

### **Example 20**

#### **Expression of pullulanase in corn**

Expression of the 6gp3 pullulanase was further analyzed by extraction from corn flour followed by PAGE and Coomassie staining. Corn-flour was made by grinding seeds, for 30 sec., in the Kleco grinder. The enzyme was extracted from about 150mg of flour with 1ml of 50mM NaOAc pH 5.5 buffer. The mixture was vortexed and incubated on a shaker at RT for 1hr, followed by another 15 min incubation at 70 °C. The mixture was then spun down (14000 rpm for 15 min at RT) and the supernatant was used as SDS-PAGE analysis. A protein band of the appropriate molecular weight (95 kDal) was observed. These samples are subjected to a pullulanase assay using commercially available dye-conjugated limit-dextrins (LIMIT-DEXTRIZYME, from Megazyme, Ireland). High levels of thermophilic pullulanase activity correlated with the presence of the 95 kD protein.

The Western blot and ELISA analysis of the transgenic corn seed also demonstrated the expression of ~95 kD protein that reacted with antibody produced against the pullulanase (expressed in *E. coli*).

### **Example 21**

#### **Increase in the rate of starch hydrolysis and improved yield of small chain (fermentable) oligosaccharides by the addition of pullulanase expressing corn**

The data shown in Figures 11A and 11B was generated from HPLC analysis, as described above, of the starch hydrolysis products from two reaction mixtures. The first reaction indicated as 'Amylase' contains a mixture [1:1 (w/w)] of corn flour samples of  $\alpha$ -amylase expressing transgenic corn made according to the method described in Example 4, for example,



and non-transgenic corn A188; and the second reaction mixture 'Amylase + Pullulanase' contains a mixture [1:1 (w/w)] of corn flour samples of  $\alpha$ -amylase expressing transgenic corn and pullulanase expressing transgenic corn made according to the method described in Example 19. The results obtained support the benefit of use of pullulanase in combination with  $\alpha$ -amylase during the starch hydrolysis processes. The benefits are from the increased rate of starch hydrolysis (Figure 11A) and increase yield of fermentable oligosaccharides with low DP (Figure 11B).

It was found that  $\alpha$ -amylase alone or  $\alpha$ -amylase and pullulanase (or any other combination of starch hydrolytic enzymes) expressed in corn can be used to produce maltodextrin (straight or branched oligosaccharides) (Figures 11A, 11B, 12, and 13A). Depending on the reaction conditions, the type of hydrolytic enzymes and their combinations, and the type of starch used the composition of the maltodextrins produced, and hence their properties, will vary.

Figure 12 depicts the results of an experiment carried out in a similar manner as described for Figure 11. The different temperature and time schemes followed during incubation of the reactions are indicated in the figure. The optimum reaction temperature for pullulanase is 75 °C and for  $\alpha$ -amylase it is >95 °C. Hence, the indicated schemes were followed to provide scope to carry out catalysis by the pullulanase and/or the  $\alpha$ -amylase at their respective optimum reaction temperature. It can be clearly deduced from the result shown that combination of  $\alpha$ -amylase and pullulanase performed better in hydrolyzing cornstarch at the end of 60 min incubation period.

HPLC analysis, as described above (except ~150 mg of corn flour was used in these reactions), of the starch hydrolysis product from two sets of reaction mixtures at the end of 30 min incubation is shown in Figure 13A and 13B. The first set of reactions was incubated at 85 °C and the second one was incubated at 95 °C. For each set there are two reaction mixtures; the first reaction indicated as 'Amylase X Pullulanase' contains flour from transgenic corn (generated by cross pollination) expressing both the  $\alpha$ -amylase and the pullulanase, and the second reaction indicated as 'Amylase' mixture of corn flour samples of  $\alpha$ -amylase expressing transgenic corn and non-transgenic corn A188 in a ratio so as to obtain same amount of  $\alpha$ -amylase activity as is observed in the cross (Amylase X Pullulanase). The total yield of low DP oligosaccharides was

more in case of  $\alpha$ -amylase and pullulanase cross compared to corn expressing  $\alpha$ -amylase alone, when the corn flour samples were incubated at 85 °C. The incubation temperature of 95 °C inactivates (at least partially) the pullulanase enzyme, hence little difference can be observed between 'Amylase X Pullulanase' and 'Amylase'. However, the data for both the incubation temperatures shows significant improvement in the amount of glucose produced (Figure 13B), at the end of the incubation period, when corn flour of  $\alpha$ -amylase and pullulanase cross was used compared to corn expressing  $\alpha$ -amylase alone. Hence use of corn expressing both  $\alpha$ -amylase and pullulanase can be especially beneficial for the processes where complete hydrolysis of starch to glucose is important.

The above examples provide ample support that pullulanase expressed in corn seeds, when used in combination with  $\alpha$ -amylase, improves the starch hydrolysis process. Pullulanase enzyme activity, being  $\alpha$  1-6 linkage specific, debranches starch far more efficiently than  $\alpha$ -amylase (an  $\alpha$  1-4 linkage specific enzyme) thereby reducing the amount of branched oligosaccharides (*e.g.* limit-dextrin, panose; these are usually non-fermentable) and increasing the amount of straight chain short oligosaccharides (easily fermentable to ethanol *etc.*). Secondly, fragmentation of starch molecules by pullulanase catalyzed debranching increases substrate accessibility for the  $\alpha$ -amylase, hence an increase in the efficiency of the  $\alpha$ -amylase catalyzed reaction results.

### Example 22

To determine whether the 797GL3 alpha amylase and malA alpha-glucosidase could function under similar pH and temperature conditions to generate an increased amount of glucose over that produced by either enzyme alone, approximately 0.35 ug of malA alpha glucosidase enzyme (produced in bacteria) was added to a solution containing 1% starch and starch purified from either non-transgenic corn seed (control) or 797GL3 transgenic corn seed (in 797GL3 corn seed the alpha amylase co-purifies with the starch). In addition, the purified starch from non-transgenic and 797GL3 transgenic corn seed was added to 1% corn starch in the absence of any malA enzyme. The mixtures were incubated at 90°C, pH 6.0 for 1 hour, spun down to remove any insoluble material, and the soluble fraction was analyzed by HPLC for glucose levels. As shown in Figure 14, the 797GL3 alpha-amylase and malA alpha-glucosidase

function at a similar pH and temperature to break down starch into glucose. The amount of glucose generated is significantly higher than that produced by either enzyme alone.

### Example 23

The utility of the *Thermoanaerobacterium* glucoamylase for raw starch hydrolysis was determined. As set forth in Figure 15, the hydrolysis conversion of raw starch was tested with water, barley  $\alpha$ -amylase (commercial preparation from Sigma), *Thermoanaerobacterium* glucoamylase, and combinations thereof were ascertained at room temperature and at 30°C. As shown, the combination of the barley  $\alpha$ -amylase with the *Thermoanaerobacterium* glucoamylase was able to hydrolyze raw starch into glucose. Moreover, the amount of glucose produced by the barley amylase and thermoanaerobacter GA is significantly higher than that produced by either enzyme alone.

### Example 24

#### Maize-optimized genes and sequences for raw-starch hydrolysis and vectors for plant transformation

The enzymes were selected based on their ability to hydrolyze raw-starch at temperatures ranging from approximately 20°-50°C. The corresponding genes or gene fragments were then designed by using maize preferred codons for the construction of synthetic genes as set forth in Example 1.

*Aspergillus shirousami*  $\alpha$ -amylase/glucoamylase fusion polypeptide (without signal sequence) was selected and has the amino acid sequence as set forth in SEQ ID NO: 45 as identified in Biosci. Biotech. Biochem., 56:884-889 (1992); Agric. Biol. Chem. 545:1905-14 (1990); Biosci. Biotechnol. Biochem. 56:174-79 (1992). The maize-optimized nucleic acid was designed and is represented in SEQ ID NO:46.

Similarly, *Thermoanaerobacterium thermosaccharolyticum* glucoamylase was selected, having the amino acid of SEQ ID NO:47 as published in Biosci. Biotech. Biochem., 62:302-308 (1998), was selected. The maize-optimized nucleic acid was designed (SEQ ID NO: 48).

*Rhizopus oryzae* glucoamylase was selected having the amino acid sequence (without signal sequence)(SEQ ID NO: 50), as described in the literature (Agric. Biol. Chem. (1986) 50, pg 957-964). The maize-optimized nucleic acid was designed and is represented in SEQ ID NO:51.

Moreover, the maize  $\alpha$ -amylase was selected and the amino acid sequence (SEQ ID NO: 51) and nucleic acid sequence (SEQ ID NO:52) were obtained from the literature. See, e.g., Plant Physiol. 105:759-760 (1994).

Expression cassettes are constructed to express the *Aspergillus shirousami*  $\alpha$ -amylase/glucoamylase fusion polypeptide from the maize-optimized nucleic acid was designed as represented in SEQ ID NO:46, the *Thermoanaerobacterium thermosaccharolyticum* glucoamylase from the maize-optimized nucleic acid was designed as represented in SEQ ID NO: 48, the *Rhizopus oryzae* glucoamylase was selected having the amino acid sequence (without signal sequence)(SEQ ID NO: 49) from the maize-optimized nucleic acid was designed and is represented in SEQ ID NO:50, and the maize  $\alpha$ -amylase.

A plasmid comprising the maize  $\gamma$ -zein N-terminal signal sequence (MRVLLVALALLALAASATS)(SEQ ID NO:17) is fused to the synthetic gene encoding the enzyme. Optionally, the sequence SEKDEL is fused to the C-terminal of the synthetic gene for targeting to and retention in the ER. The fusion is cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm in a plant transformation plasmid. The fusion is delivered to the corn tissue via *Agrobacterium* transfection.

#### Example 25

Expression cassettes comprising the selected enzymes are constructed to express the enzymes. A plasmid comprising the sequence for a raw starch binding site is fused to the synthetic gene encoding the enzyme. The raw starch binding site allows the enzyme fusion to bind to non-gelatinized starch. The raw-starch binding site amino acid sequence (SEQ ID NO:53) was determined based on literature, and the nucleic acid sequence was maize-optimized to give SEQ ID NO:54. The maize-optimized nucleic acid sequence is fused to the synthetic gene encoding the enzyme in a plasmid for expression in a plant.

### Example 26

#### Construction of maize-optimized genes and vectors for plant transformation

The genes or gene fragments were designed by using maize preferred codons for the construction of synthetic genes as set forth in Example 1.

*Pyrococcus furiosus* EGLA, hyperthermophilic endoglucanase amino acid sequence (without signal sequence) was selected and has the amino acid sequence as set forth in SEQ ID NO: 55, as identified in Journal of Bacteriology (1999) 181, pg 284-290.) The maize-optimized nucleic acid was designed and is represented in SEQ ID NO:56.

*Thermus flavus* xylose isomerase was selected and has the amino acid sequence as set forth in SEQ ID NO:57, as described in Applied Biochemistry and Biotechnology 62:15-27 (1997).

Expression cassettes are constructed to express the *Pyrococcus furiosus* EGLA (endoglucanase) from the maize-optimized nucleic acid (SEQ ID NO:56) and the *Thermus flavus* xylose isomerase from a maize-optimized nucleic acid encoding amino acid sequence SEQ ID NO:57. A plasmid comprising the maize  $\gamma$ -zein N-terminal signal sequence (MRVLLVALALLALAASATS)(SEQ ID NO:17) is fused to the synthetic maize-optimized gene encoding the enzyme. Optionally, the sequence SEKDEL is fused to the C-terminal of the synthetic gene for targeting to and retention in the ER. The fusion is cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm in a plant transformation plasmid. The fusion is delivered to the corn tissue via *Agrobacterium* transfection.

### Example 27

#### Production of glucose from corn flour using thermophilic enzymes expressed in corn

Expression of the hyperthermophilic  $\alpha$ -amylase, 797GL3 and  $\alpha$ -glucosidase (Mala) were shown to result in production of glucose when mixed with an aqueous solution and incubated at 90 °C

A transgenic corn line (line 168A10B, pNOV4831) expressing Mala enzyme was identified by measuring  $\alpha$ -glucosidase activity as indicated by hydrolysis of p-nitrophenyl- $\alpha$ -glucoside.

Corn kernels from transgenic plants expressing 797GL3 were ground to a flour in a Kleco cell thus creating amylase flour. Corn kernels from transgenic plants expressing MalA were ground to a flour in a Kleco cell thus creating MalA flour. Non-transgenic corn kernels were ground in the same manner to generate control flour.

Buffer was 50 mM MES buffer pH 6.0.

Corn flour hydrolysis reactions: Samples were prepared as indicated in Table 5 below. Corn flour (about 60 mg per sample) was mixed with 40 ml of 50 mM MES buffer, pH 6.0. Samples were incubated in a water bath set at 90°C for 2.5 and 14 hours. At the indicated incubation times, samples were removed and analyzed for glucose content.

The samples were assayed for glucose by a glucose oxidase / horse radish peroxidase based assay. GOPOD reagent contained: 0.2 mg/ml o-dianisidine, 100 mM Tris pH 7.5, 100 U/ml glucose oxidase & 10 U/ml horse radish peroxidase. 20 µl of sample or diluted sample were arrayed in a 96 well plate along with glucose standards (which varied from 0 to 0.22 mg/ml). 100 µl of GOPOD reagent was added to each well with mixing and the plate incubated at 37 °C for 30 min. 100 µl of sulfuric acid (9M) was added and absorbance at 540 nm was read. The glucose concentration of the samples was determined by reference to the standard curve. The quantity of glucose observed in each sample is indicated in Table 5.

Table 5

Sample	WT flour mg	amylase flour mg	MaIA flour Mg	Buffer ml	Glucose 2.5 h mg	Glucose 14 h mg
1	66	0	0	40	0	0
2	31	30	0	40	0.26	0.50
3	30	0	31.5	40	0	0.09
4	0	32.2	30.0	40	2.29	12.30
5	0	6.1	56.2	40	1.16	8.52

These data demonstrate that when expression of hyperthermophilic  $\alpha$ -amylase and  $\alpha$ -glucosidase in corn result in a corn product that will generate glucose when hydrated and heated under appropriate conditions.

### Example 28

#### Production of Maltodextrins

Grain expressing thermophilic  $\alpha$ -amylase was used to prepare maltodextrins. The exemplified process does not require prior isolation of the starch nor does it require addition of exogenous enzymes.

Corn kernels from transgenic plants expressing 797GL3 were ground to a flour in a Kleco cell to create "amylase flour". A mixture of 10% transgenic/90% non-transgenic kernels was ground in the same manner to create "10% amylase flour."

Amylase flour and 10% amylase flour (approximately 60 mg/sample) were mixed with water at a rate of 5  $\mu$ l of water per mg of flour. The resulting slurries were incubated at 90°C for up to 20 hours as indicated in Table 6. Reactions were stopped by addition of 0.9 ml of 50 mM EDTA at 85°C and mixed by pipetting. Samples of 0.2 ml of slurry were removed, centrifuged to remove insoluble material and diluted 3x in water.

The samples were analyzed by HPLC with ELSD detection for sugars and maltodextrins. The gradient HPLC system was equipped with Astec Polymer Amino Column, 5 micron particle size,

250 X 4.6 mm and an Alltech ELSD 2000 detector. The system was pre-equilibrated with a 15:85 mixture of water:acetonitrile. The flow rate was 1 ml/min. The initial conditions were maintained for 5 min after injection followed by a 20 min gradient to 50:50 water:acetonitrile followed by 10 minutes of the same solvent. The system was washed with 20 min of 80:20 water:acetonitrile and then re-equilibrated with the starting solvent.

The resulting peak areas were normalized for volume and weight of flour. The response factor of ELSD per  $\mu\text{g}$  of carbohydrate decreases with increasing DP, thus the higher DP maltodextrins represent a higher percentage of the total than indicated by peak area.

The relative peak areas of the products of reactions with 100% amylase flour are shown in Figure 17. The relative peak areas of the products of reactions with 10% amylase flour are shown in Figure 18.

These data demonstrate that a variety of maltodextrin mixtures can be produced by varying the time of heating. The level of  $\alpha$ -amylase activity can be varied by mixing transgenic  $\alpha$ -amylase-expressing corn with wild-type corn to alter the maltodextrin profile.

The products of the hydrolysis reactions described in this example can be concentrated and purified for food and other applications by use of a variety of well defined methods including: centrifugation, filtration, ion-exchange, gel permeation, ultrafiltration, nanofiltration, reverse osmosis, decolorizing with carbon particles, spray drying and other standard techniques known to the art.

## Example 29

### Effect of time and temperature on maltodextrin production

The composition of the maltodextrin products of autohydrolysis of grain containing thermophilic  $\alpha$ -amylase may be altered by varying the time and temperature of the reaction.

In another experiment, amylase flour was produced as described in Example 28 above and mixed with water at a ratio of 300 $\mu\text{l}$  water per 60 mg flour. Samples were incubated at 70°, 80°, 90°, or 100° C for up to 90 minutes. Reactions were stopped by addition of 900ml of 50mM EDTA at 90°C, centrifuged to remove insoluble material and filtered through 0.45 $\mu\text{m}$  nylon filters. Filtrates were analyzed by HPLC as described in Example 28.



The result of this analysis is presented in Figure 19. The DP number nomenclature refers to the degree of polymerization. DP2 is maltose; DP3 is maltotriose, etc. Larger DP maltodextrins eluted in a single peak near the end of the elution and are labeled ">DP12". This aggregate includes dextrans that passed through 0.45  $\mu$ m filters and through the guard column and does not include any very large starch fragments trapped by the filter or guard column.

This experiment demonstrates that the maltodextrin composition of the product can be altered by varying both temperature and incubation time to obtain the desired maltooligosaccharide or maltodextrin product.

### Example 30

#### Maltodextrin production

The composition of maltodextrin products from transgenic maize containing thermophilic  $\alpha$ -amylase can also be altered by the addition of other enzymes such as  $\alpha$ -glucosidase and xylose isomerase as well as by including salts in the aqueous flour mixture prior to treating with heat.

In another, amylase flour, prepared as described above, was mixed with purified MalA and/or a bacterial xylose isomerase, designated BD8037. *S. sulfotaricus* MalA with a 6His purification tag was expressed in *E. coli*. Cell lysate was prepared as described in Example 28, then purified to apparent homogeneity using a nickel affinity resin (Probond, Invitrogen) and following the manufacturer's instructions for native protein purification. Xylose isomerase BD8037 was obtained as a lyophilized powder from Diversa and resuspended in 0.4x the original volume of water.

Amylase corn flour was mixed with enzyme solutions plus water or buffer. All reactions contained 60 mg amylase flour and a total of 600 $\mu$ l of liquid. One set of reactions was buffered with 50 mM MOPS, pH 7.0 at room temperature, plus 10mM MgSO<sub>4</sub> and 1 mM CoCl<sub>2</sub>; in a second set of reactions the metal-containing buffer solution was replaced by water. All reactions were incubated for 2 hours at 90°C. Reaction supernatant fractions were prepared by centrifugation. The pellets were washed with an additional 600 $\mu$ l H<sub>2</sub>O and re-centrifuged. The supernatant fractions from each reaction were combined, filtered through a Centricon 10, and analyzed by HPLC with ELSD detection as described above.

The results are graphed in Figure 20. They demonstrate that the grain-expressed amylase 797GL3 can function with other thermophilic enzymes, with or without added metal ions, to produce a variety of maltodextrin mixtures from corn flour at a high temperature. In particular, the inclusion of a glucoamylase or  $\alpha$ -glucosidase may result in a product with more glucose and other low DP products. Inclusion of an enzyme with glucose isomerase activity results in a product that has fructose and thus would be sweeter than that produced by amylase alone or amylase with  $\alpha$ -glucosidase. In addition the data indicate that the proportion of DP5, DP6 and DP7 maltooligosaccharides can be increased by including divalent cationic salts, such as  $\text{CoCl}_2$  and  $\text{MgSO}_4$ .

Other means of altering the maltodextrin composition produced by a reaction such as that described here include: varying the reaction pH, varying the starch type in the transgenic or non-transgenic grain, varying the solids ratio, or by addition of organic solvents.

### **Example 31**

#### Preparing dextrins, or sugars from grain without mechanical disruption of the grain prior to recovery of starch-derived products

Sugars and maltodextrins were prepared by contacting the transgenic grain expressing the  $\alpha$ -amylase, 797GL3, with water and heating to 90°C overnight (>14 hours). Then the liquid was separated from the grain by filtration. The liquid product was analyzed by HPLC by the method described in Example 15. Table 6 presents the profile of products detected.

Table 6

Molecular species	Concentration of Products $\mu\text{g} / 25 \mu\text{l}$ injection
Fructose	0.4
Glucose	18.0
Maltose	56.0
DP3*	26.0
DP4*	15.9
DP5*	11.3
DP6*	5.3
DP7*	1.5

\* Quantification of DP3 includes maltotriose and may include isomers of maltotriose that have an  $\alpha(1\rightarrow6)$  bond in place of an  $\alpha(1\rightarrow4)$  bond. Similarly DP4 to DP7 quantification includes the linear maltooligosaccharides of a given chain length as well as isomers that have one or more  $\alpha(1\rightarrow6)$  bonds in place of one or more  $\alpha(1\rightarrow4)$  bonds

These data demonstrate that sugars and maltodextrins can be prepared by contacting intact  $\alpha$ -amylase-expressing grain with water and heating. The products can then be separated from the intact grain by filtration or centrifugation or by gravitational settling.

### Example 32

#### Fermentation of raw starch in corn expressing *Rhizopus oryzae* glucoamylase.

Transgenic corn kernels are harvested from transgenic plants made as described in Example 29. The kernels are ground to a flour. The corn kernels express a protein that contains an active fragment of the glucoamylase of *Rhizopus oryzae* (Sequence ID NO: 49) targeted to the endoplasmic reticulum.

The corn kernels are ground to a flour as described in Example 15. Then a mash is prepared containing s 20 g of corn flour, 23 ml of de-ionized water, 6.0 ml of backset (8% solids by weight). pH is adjusted to 6.0 by addition of ammonium hydroxide. The following

components are added to the mash: protease (0.60 ml of a 1,000-fold dilution of a commercially available protease), 0.2 mg Lactocide & urea (0.85 ml of a 10-fold dilution of 50% Urea Liquor). A hole is cut into the cap of the 100 ml bottle containing the mash to allow CO<sub>2</sub> to vent. The mash is then inoculated with yeast (1.44 ml) and incubated in a water bath set at 90° C. After 24 hours of fermentation the temperature is lowered to 86°C; at 48 hours it is set to 82 °C.

Yeast for inoculation is propagated as described in Example 14.

Samples are removed as described in example 14 and then analyzed by the methods described in Example 14.

### Example 33

Transgenic corn kernels are harvested from transgenic plants made as described in Example 28. The kernels are ground to a flour. The corn kernels express a protein that contains an active fragment of the glucoamylase of *Rhizopus oryzae* (Sequence ID NO: 49) targeted to the endoplasmic reticulum.

The corn kernels are ground to a flour as described in Example 15. Then a mash is prepared containing 20 g of corn flour, 23 ml of de-ionized water, 6.0 ml of backset (8% solids by weight). pH is adjusted to 6.0 by addition of ammonium hydroxide. The following components are added to the mash: protease (0.60 ml of a 1,000-fold dilution of a commercially available protease), 0.2 mg Lactocide & urea (0.85 ml of a 10-fold dilution of 50% Urea Liquor). A hole is cut into the cap of the 100 ml bottle containing the mash to allow CO<sub>2</sub> to vent. The mash is then inoculated with yeast (1.44 ml) and incubated in a water bath set at 90° C. After 24 hours of fermentation the temperature is lowered to 86° C; at 48 hours it is set to 82° C.

Yeast for inoculation is propagated as described in Example 14.

Samples are removed as described in example 14 and then analyzed by the methods described in Example 14.

### Example 34

Example of fermentation of raw starch in whole kernels of corn expressing  
*Rhizopus oryzae* glucoamylase with addition of exogenous  $\alpha$ -amylase

Transgenic corn kernels are harvested from transgenic plants made as described in Example 28. The corn kernels express a protein that contains an active fragment of the glucoamylase of *Rhizopus oryzae* (Sequence ID NO: 49) targeted to the endoplasmic reticulum.

The corn kernels are contacted with 20 g of corn flour, 23 ml of de-ionized water, 6.0 ml of backset (8% solids by weight). pH is adjusted to 6.0 by addition of ammonium hydroxide. The following components are added: barley  $\alpha$ -amylase purchased from Sigma (2 mg), protease (0.60 ml of a 1,000-fold dilution of a commercially available protease), 0.2 mg Lactocide & urea (0.85 ml of a 10-fold dilution of 50% Urea Liquor). A hole is cut into the cap of the 100 ml bottle containing the mixture in order to allow CO<sub>2</sub> to vent. The mixture is then inoculated with yeast (1.44 ml) and incubated in a water bath set at 90° C. After 24 hours of fermentation the temperature is lowered to 86° C; at 48 hours it is set to 82° C.

Yeast for inoculation is propagated as described in Example 14.

Samples are removed as described in example 14 and then analyzed by the methods described in Example 14.

### Example 35

#### Fermentation of raw starch in corn expressing *Rhizopus oryzae* glucoamylase and *Zea mays* amylase

Transgenic corn kernels are harvested from transgenic plants made as described in Example 28. The corn kernels express a protein that contains an active fragment of the glucoamylase of *Rhizopus oryzae* (Sequence ID NO:49) targeted to the endoplasmic reticulum. The kernels also express the maize amylase with raw starch binding domain as described in Example 28.

The corn kernels are ground to a flour as described in Example 14. Then a mash is prepared containing 20 g of corn flour, 23 ml of de-ionized water, 6.0 ml of backset (8% solids by weight). pH is adjusted to 6.0 by addition of ammonium hydroxide. The following components are added to the mash: protease (0.60 ml of a 1,000-fold dilution of a commercially available protease), 0.2 mg Lactocide & urea (0.85 ml of a 10-fold dilution of 50% Urea Liquor). A hole is cut into the cap of the 100 ml bottle containing the mash to allow CO<sub>2</sub> to vent. The

mash is then inoculated with yeast (1.44 ml) and incubated in a water bath set at 90 F. After 24 hours of fermentation the temperature is lowered to 86 F; at 48 hours it is set to 82 F.

Yeast for inoculation is propagated as described in Example 14.

Samples are removed as described in example 14 and then analyzed by the methods described in Example 14.

### Example 36

#### Example of fermentation of raw starch in corn expressing

#### *Thermoanaerobacter thermosaccharolyticum* glucoamylase.

Transgenic corn kernels are harvested from transgenic plants made as described in Example 28. The corn kernels express a protein that contains an active fragment of the glucoamylase of *Thermoanaerobacter thermosaccharolyticum* (Sequence ID NO: 47) targeted to the endoplasmic reticulum.

The corn kernels are ground to a flour as described in Example 15. Then a mash is prepared containing 20 g of corn flour, 23 ml of de-ionized water, 6.0 ml of backset (8% solids by weight). pH is adjusted to 6.0 by addition of ammonium hydroxide. The following components are added to the mash: protease (0.60 ml of a 1,000-fold dilution of a commercially available protease), 0.2 mg Lactocide & urea (0.85 ml of a 10-fold dilution of 50% Urea Liquor). A hole is cut into the cap of the 100 ml bottle containing the mash to allow CO<sub>2</sub> to vent. The mash is then inoculated with yeast (1.44 ml) and incubated in a water bath set at 90° C. After 24 hours of fermentation the temperature is lowered to 86° C; at 48 hours it is set to 82° C.

Yeast for inoculation is propagated as described in Example 14.

Samples are removed as described in example 14 and then analyzed by the methods described in Example 14.

### Example 37

#### Example of fermentation of raw starch in corn expressing

#### *Aspergillus niger* glucoamylase

Transgenic corn kernels are harvested from transgenic plants made as described in Example 28. The corn kernels express a protein that contains an active fragment of the glucoamylase of *Aspergillus niger* (Fiil,N.P. "Glucoamylases G1 and G2 from *Aspergillus niger*

are synthesized from two different but closely related mRNAs" EMBO J. 3 (5), 1097-1102 (1984), Accession number P04064). The maize-optimized nucleic acid encoding the glucoamylase has SEQ ID NO:59 and is targeted to the endoplasmic reticulum.

The corn kernels are ground to a flour as described in Example 14. Then a mash is prepared containing 20 g of corn flour, 23 ml of de-ionized water, 6.0 ml of backset (8% solids by weight). pH is adjusted to 6.0 by addition of ammonium hydroxide. The following components are added to the mash: protease (0.60 ml of a 1,000-fold dilution of a commercially available protease), 0.2 mg Lactocide & urea (0.85 ml of a 10-fold dilution of 50% Urea Liquor). A hole is cut into the cap of the 100 ml bottle containing the mash to allow CO<sub>2</sub> to vent. The mash is then inoculated with yeast (1.44 ml) and incubated in a water bath set at 90° C. After 24 hours of fermentation the temperature is lowered to 86° C; at 48 hours it is set to 82° C.

Yeast for inoculation is propagated as described in Example 14.

Samples are removed as described in example 14 and then analyzed by the methods described in Example 14.

### Example 38

#### Example of fermentation of raw starch in corn expressing

#### *Aspergillus niger* glucoamylase and *Zea mays* amylase

Transgenic corn kernels are harvested from transgenic plants made as described in Example 28. The corn kernels express a protein that contains an active fragment of the glucoamylase of *Aspergillus niger* (Fiil,N.P. "Glucoamylases G1 and G2 from *Aspergillus niger* are synthesized from two different but closely related mRNAs" EMBO J. 3 (5), 1097-1102 (1984) : Accession number P04064)(SEQ ID NO:59, maize-optimized nucleic acid) and is targeted to the endoplasmic reticulum. The kernels also express the maize amylase with raw starch binding domain as described in example 28.

The corn kernels are ground to a flour as described in Example 14. Then a mash is prepared containing 20 g of corn flour, 23 ml of de-ionized water, 6.0 ml of backset (8% solids by weight). pH is adjusted to 6.0 by addition of ammonium hydroxide. The following components are added to the mash: protease (0.60 ml of a 1,000-fold dilution of a commercially available protease), 0.2 mg Lactocide & urea (0.85 ml of a 10-fold dilution of 50% Urea Liquor).

A hole is cut into the cap of the 100 ml bottle containing the mash to allow CO<sub>2</sub> to vent. The mash is then inoculated with yeast (1.44 ml) and incubated in a water bath set at 90° C. After 24 hours of fermentation the temperature is lowered to 86° C; at 48 hours it is set to 82° C.

Yeast for inoculation is propagated as described in Example 14.

Samples are removed as described in example 14 and then analyzed by the methods described in Example 14.

### Example 39

#### Example of fermentation of raw starch in corn expressing

##### *Thermoanaerobacter thermosaccharolyticum* glucoamylase and barley amylase

Transgenic corn kernels are harvested from transgenic plants made as described in Example 28. The corn kernels express a protein that contains an active fragment of the glucoamylase of *Thermoanaerobacter thermosaccharolyticum* (Sequence ID NO: 47) targeted to the endoplasmic reticulum. The kernels also express the low pI barley amylase *amyl* gene (Rogers, J.C. and Milliman, C. "Isolation and sequence analysis of a barley alpha-amylase cDNA clone" J. Biol. Chem. 258 (13), 8169-8174 (1983) modified to target expression of the protein to the endoplasmic reticulum.

The corn kernels are ground to a flour as described in Example 14. Then a mash is prepared containing 20 g of corn flour, 23 ml of de-ionized water, 6.0 ml of backset (8% solids by weight). pH is adjusted to 6.0 by addition of ammonium hydroxide. The following components are added to the mash: protease (0.60 ml of a 1,000-fold dilution of a commercially available protease), 0.2 mg Lactocid & urea (0.85 ml of a 10-fold dilution of 50% Urea Liquor). A hole is cut into the cap of the 100 ml bottle containing the mash to allow CO<sub>2</sub> to vent. The mash is then inoculated with yeast (1.44 ml) and incubated in a water bath set at 90° C. After 24 hours of fermentation the temperature is lowered to 86° C; at 48 hours it is set to 82° C.

Yeast for inoculation is propagated as described in Example 14.

Samples are removed as described in example 14 and then analyzed by the methods described in Example 14.



#### Example 40

##### Example of fermentation of raw starch in whole kernels of corn expressing *Thermoanaerobacter thermosaccharolyticum* glucoamylase and barley amylase.

Transgenic corn kernels are harvested from transgenic plants made as described in Example 28. The corn kernels express a protein that contains an active fragment of the glucoamylase of *Thermoanaerobacter thermosaccharolyticum* (Sequence ID NO: 47) targeted to the endoplasmic reticulum. The kernels also express the low pI barley amylase amy1 gene (Rogers, J.C. and Milliman, C. "Isolation and sequence analysis of a barley alpha-amylase cDNA clone" J. Biol. Chem. 258 (13), 8169-8174 (1983) modified to target expression of the protein to the endoplasmic reticulum.

The corn kernels are contacted with 20 g of corn flour, 23 ml of de-ionized water, 6.0 ml of backset (8% solids by weight). pH is adjusted to 6.0 by addition of ammonium hydroxide. The following components are added to the mixture: protease (0.60 ml of a 1,000-fold dilution of a commercially available protease), 0.2 mg Lactocide & urea (0.85 ml of a 10-fold dilution of 50% Urea Liquor). A hole is cut into the cap of the 100 ml bottle containing the mash to allow CO<sub>2</sub> to vent. The mixture is then inoculated with yeast (1.44 ml) and incubated in a water bath set at 90° C. After 24 hours of fermentation the temperature is lowered to 86° C; at 48 hours it is set to 82° C.

Yeast for inoculation is propagated as described in Example 14.

Samples are removed as described in example 14 and then analyzed by the methods described in Example 14.

#### Example 41

##### Example of fermentation of raw starch in corn expressing an alpha-amylase and glucoamylase fusion.

Transgenic corn kernels are harvested from transgenic plants made as described in Example 28. The corn kernels express a maize-optimized polynucleotide such as provided in SEQ ID NO: 46, encoding an alpha-amylase and glucoamylase fusion, such as provided in SEQ ID NO: 45, which are targeted to the endoplasmic reticulum. .

The corn kernels are ground to a flour as described in Example 14. Then a mash is prepared containing 20 g of corn flour, 23 ml of de-ionized water, 6.0 ml of backset (8% solids

by weight). pH is adjusted to 6.0 by addition of ammonium hydroxide. The following components are added to the mash: protease (0.60 ml of a 1,000-fold dilution of a commercially available protease), 0.2 mg Lactocide & urea (0.85 ml of a 10-fold dilution of 50% Urea Liquor). A hole is cut into the cap of the 100 ml bottle containing the mash to allow CO<sub>2</sub> to vent. The mash is then inoculated with yeast (1.44 ml) and incubated in a water bath set at 90° C. After 24 hours of fermentation the temperature is lowered to 86° C; at 48 hours it is set to 82° C.

Yeast for inoculation is propagated as described in Example 14.

Samples are removed as described in example 14 and then analyzed by the methods described in Example 14.

### **Example 42**

#### **Construction of transformation vectors**

Expression cassettes were constructed to express the hyperthermophilic beta-glucanase EglA in maize as follows:

pNOV4800 comprises the barley Amy32b signal peptide (MGKNGNLCCFSLLLLLLAGLASGHQ) fused to the synthetic gene for the EglA beta-glucanase for targeting to the endoplasmic reticulum and secretion into the apoplast. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

pNOV4803 comprises the barley Amy32b signal peptide fused to the synthetic gene for the EglA beta-glucanase for targeting to the endoplasmic reticulum and secretion into the apoplast. The fusion was cloned behind the maize ubiquitin promoter for expression throughout the plant.

Expression cassettes were constructed to express the thermophilic beta-glucanase/mannanase 6GP1 (SEQ ID NO: 85) in maize as follows:

**pNOV4819** comprises the tobacco PR1a signal peptide (MGFVLFSQLPSFLLVSTLLLFLVISHSCRA) fused to the synthetic gene for the 6GP1 beta-glucanase/mannanase for targeting to the endoplasmic reticulum and secretion into the apoplast. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

**pNOV4820** comprises the synthetic gene for 6GP1 cloned behind the maize  $\gamma$ -zein promoter for cytoplasmic localization and expression specifically in the endosperm.

**pNOV4823** comprises the tobacco PR1a signal peptide fused to the synthetic gene for the 6GP1 beta-glucanase/mannanase with a C-terminal addition of the sequence KDEL for targeting to and retention in the endoplasmic reticulum. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

**pNOV4825** comprises the tobacco PR1a signal peptide fused to the synthetic gene for the 6GP1 beta-glucanase/mannanase with a C-terminal addition of the sequence KDEL for targeting to and retention in the endoplasmic reticulum. The fusion was cloned behind the maize ubiquitin promoter for expression throughout the plant.

Expression cassettes were constructed to express the barley Amyl alpha-amylase (SEQ ID NO: 87) in maize as follows:

**pNOV4867** comprises the maize  $\gamma$ -zein N-terminal signal sequence fused to the barley Amyl alpha-amylase with a C-terminal addition of the sequence SEKDEL for targeting to and retention in the endoplasmic reticulum. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

**pNOV4879** comprises the maize  $\gamma$ -zein N-terminal signal sequence fused to the barley Amyl alpha-amylase with a C-terminal addition of the sequence SEKDEL for targeting to and retention

in the endoplasmic reticulum. The fusion was cloned behind the maize globulin promoter for expression specifically in the embryo.

**pNOV4897** comprises the maize  $\gamma$ -zein N-terminal signal sequence fused to the barley AmyI alpha-amylase for targeting to the endoplasmic reticulum and secretion into the apoplast. The fusion was cloned behind the maize globulin promoter for expression specifically in the embryo.

**pNOV4895** comprises the maize  $\gamma$ -zein N-terminal signal sequence fused to the barley AmyI alpha-amylase for targeting to the endoplasmic reticulum and secretion into the apoplast. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

**pNOV4901** comprises the gene for the barley AmyI alpha-amylase cloned behind the maize globulin promoter for cytoplasmic localization and expression specifically in the embryo.

Expression cassettes were constructed to express the *Rhizopus* glucoamylase (SEQ ID NO: 50) in maize as follows:

**pNOV4872** comprises the maize  $\gamma$ -zein N-terminal signal sequence fused to the synthetic gene for *Rhizopus* glucoamylase with a C-terminal addition of the sequence SEKDEL for targeting to and retention in the endoplasmic reticulum. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

**pNOV4880** comprises the maize  $\gamma$ -zein N-terminal signal sequence fused to the synthetic gene for *Rhizopus* glucoamylase with a C-terminal addition of the sequence SEKDEL for targeting to and retention in the endoplasmic reticulum. The fusion was cloned behind the maize globulin promoter for expression specifically in the embryo.

**pNOV4889** comprises the maize  $\gamma$ -zein N-terminal signal sequence fused to the synthetic gene for *Rhizopus* glucoamylase for targeting to the endoplasmic reticulum and secretion into the

apoplast. The fusion was cloned behind the maize globulin promoter for expression specifically in the embryo.

**pNOV4890** comprises the maize  $\gamma$ -zein N-terminal signal sequence fused to the synthetic gene for *Rhizopus* glucoamylase for targeting to the endoplasmic reticulum and secretion into the apoplast. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

**pNOV4891** comprises the synthetic gene for *Rhizopus* glucoamylase cloned behind the maize  $\gamma$ -zein promoter for cytoplasmic localization and expression specifically in the endosperm.

### **Example 43**

#### Expression of the mesophilic *Rhizopus* glucoamylase in corn

A variety of constructs were generated for the expression of the *Rhizopus* glucoamylase in corn. The maize  $\gamma$ -zein and globulin promoters were used to express the glucoamylase specifically in the endosperm or embryo, respectively. In addition, the maize  $\gamma$ -zein signal sequence and a synthetic ER retention signal were used to regulate the subcellular localization of the glucoamylase protein. All 5 constructs (pNOV4872, pNOV4880, pNOV4889, pNOV4890, and pNOV4891) yielded transgenic plants with glucoamylase activity detected in the seed. Tables 7 and 8 show the results for individual transgenic seed (construct pNOV4872) and pooled seed (construct pNOV4889), respectively. No detrimental phenotype was observed for any transgenic plants expressing this *Rhizopus* glucoamylase.

**Glucoamylase assay:** Seed were ground to a flour and the flour was suspended in water. The samples were incubated at 30 degrees for 50 minutes to allow the glucoamylase to react with the starch. The insoluble material was pelleted and the glucose concentration was determined for the supernatants. The amount of glucose liberated in each sample was taken as an indication of the level of glucoamylase present. Glucose concentration was determined by incubating the samples with GOHOD reagent (300mM Tris/Cl pH7.5, glucose oxidase

(20U/ml), horseradish peroxidase (20U/ml), o-dianisidine 0.1 mg/ml) for 30 minutes at 37 degrees C, adding 0.5 volumes of 12N H<sub>2</sub>SO<sub>4</sub>, and measuring the OD<sub>540</sub>.

Table 7 shows activity of the *Rhizopus glucoamylase* in individual transgenic corn seed (construct pNOV4872).

Seed	U/g flour
Wild Type #1	0.07
Wild Type #2	0.55
Wild Type #3	0.25
Wild Type #4	0.33
Wild Type #5	0.30
Wild Type #6	0.42
Wild Type #7	-0.01
Wild Type #8	0.31
MD9L022156 #1	5.17
MD9L022156 #2	1.66
MD9L022156 #3	7.66
MD9L022156 #4	1.77
MD9L022156 #5	7.08
MD9L022156 #6	4.46
MD9L022156 #7	2.20
MD9L022156 #8	3.50
MD9L023377 #1	9.23
MD9L023377 #2	4.30
MD9L023377 #3	6.72
MD9L023377 #4	3.35
MD9L023377 #5	0.56
MD9L023377 #6	4.79
MD9L023377 #7	4.60
MD9L023377 #8	6.01
MD9L023043 #1	4.93
MD9L023043 #2	8.74
MD9L023043 #3	2.70
MD9L023043 #4	0.72
MD9L023043 #5	3.33
MD9L023043 #6	3.53
MD9L023043 #7	3.94
MD9L023043 #8	11.51

MD9L023334 #1	4.28
MD9L023334 #2	2.86
MD9L023334 #3	0.56
MD9L023334 #4	6.96
MD9L023334 #5	3.29
MD9L023334 #6	3.18
MD9L023334 #7	4.57
MD9L023334 #8	7.44
MD9L022039 #1	6.25
MD9L022039 #2	2.85
MD9L022039 #3	4.32
MD9L022039 #4	2.51
MD9L022039 #5	5.06
MD9L022039 #6	5.03
MD9L022039 #7	2.79
MD9L022039 #8	2.98

Table 8 shows activity of the *Rhizopus* glucoamylase in pooled transgenic corn seed (construct pNOV4889).

**Table 8**

<b>Seed</b>	<b>U/g flour</b>
Wild Type	0.38
MD9L023347	2.14
MD9L023352	2.34
MD9L023369	1.66
MD9L023469	1.42
MD9L023477	1.33
MD9L023482	1.95
MD9L023484	1.32
MD9L024170	1.35
MD9L024177	1.48
MD9L024184	1.60
MD9L024186	1.34
MD9L024196	1.38
MD9L024228	1.69
MD9L024263	1.70
MD9L024315	1.32

MD9L024325	1.73
MD9L024333	1.41
MD9L024339	1.84

All expression cassettes were inserted into the binary vector pNOV2117 for transformation into maize via *Agrobacterium* infection. The binary vector contained the phosphomannose isomerase (PMI) gene which allows for selection of transgenic cells with mannose. Transformed maize plants were either self-pollinated or outcrossed and seed was collected for analysis.

#### Example 44

##### Expression of the hyperthermophilic beta-glucanase EglA in corn

For expression of the hyperthermophilic beta-glucanase EglA in corn we utilized the ubiquitin promoter for expression throughout the plant and the  $\gamma$ -zein promoter for expression specifically in the endosperm of corn seed. The barley Amy32b signal peptide was fused to EglA for localization in the apoplast.

Expression of the hyperthermophilic beta-glucanase EglA in transgenic corn seed and leaves was analysed using an enzymatic assay and western blotting.

Transgenic seed segregating for construct pNOV4800 or pNOV4803 were analysed using both western blotting and an enzymatic assay for beta-glucanase. Endosperm was isolated from individual seed after soaking in water for 48 hours. Protein was extracted by grinding the endosperm in 50mM NaPO<sub>4</sub> buffer (pH 6.0). Heat-stable proteins were isolated by heating the extracts at 100 degrees C for 15 minutes, followed by pelleting of the insoluble material. The supernatant containing heat-stable proteins was analysed for beta glucanase activity using the azo-barley glucan method (megazyme). Samples were pre-incubated at 100 degrees C for 10 minutes and assayed for 10 minutes at 100 degrees C using the azo-barley glucan substrate. Following incubation, 3 volumes of precipitation solution were added to each sample, the samples were centrifuged for 1 minute, and the OD590 of each supernatant was determined. In addition, 5ug of protein were separated by SDS-PAGE and blotted to nitrocellulose for western



blot analysis using antibodies against the EglA protein. Western blot analysis detected a specific, heat-stable protein(s) in the EglA positive endosperm extracts, and not in negative extracts. The western blot signal correlates with the level of EglA activity detected enzymatically.

EglA activity was analysed in leaves and seed of plants containing the transgenic constructs pNOV4803 and pNOV4800, respectively. The assays (conducted as described above) showed that the heat-stable beta-glucanase EglA was expressed at various levels in the leaves (Table 9) and seed (Table 10) of transgenic plants while no activity was detected in non-transgenic control plants. Expression of EglA in corn utilizing constructs pNOV4800 and pNOV4803 did not result in any detectable negative phenotype.

Table 9 shows the activity of the hyperthermophilic beta-glucanase EglA in leaves of transgenic corn plants. Enzymatic assays were conducted on extracts from leaves of pNOV4803 transgenic plants to detect hyperthermophilic beta-glucanase activity. Assays were conducted at 100 degrees C using the azo-barley glucan method (megazyme). The results indicate that the transgenic leaves have varying levels of hyperthermophilic beta-glucanase activity.

Table 9

Plant	Abs590
Wild Type	0
266A-17D	0.008
266A-18E	0.184
266A-13C	0.067
266A-15E	0.003
266A-11E	0
265C-1B	0.024
265C-1C	0.065
265C-2D	0.145
265C-5C	0.755
265C-5D	0.133
265C-3A	0.076
266A-4B	0.045
266A-12B	0.066
266A-11C	0.096

266A-14B	0.074
266A-4C	0.107
266A-4A	0.084
266A-12A	0.054
266A-15B	0.052
266A-11A	0.109
266A-20C	0.044
266A-19D	0.02
266A-12C	0.098
266A-4E	0.248
266A-18B	0.367
265C-3D	0.066
266A-20E	0.163
266A-13D	0.084
265C-3B	0.065
266A-15A	0.131
266A-13A	0.169
265C-3E	0.116
266A-20A	0.365
266A-20B	0.521
266A-19C	0.641
266A-20D	0.561
266A-4D	0.363
266A-18A	0.676
265C-5E	0.339
266A-17E	0.221
266A-11B	0.251
265C-4E	0.138
265C-4D	0.242

Table 10 shows the activity of the hyperthermophilic beta-glucanase EglA in seed of transgenic corn plants. Enzymatic assays were conducted on extracts from individual, segregating seed of pNOV4800 transgenic plants to detect hyperthermophilic beta-glucanase activity. Assays were conducted at 100 degrees C using the azo-barley glucan method (megazyme). The results indicate that the transgenic seed have varying levels of hyperthermophilic beta-glucanase activity.

Table 10

Seed	Abs 590
Wild Type	0
1A	1.1
1B	0
1C	1.124
1D	1.323
2A	0
2B	1.354
2C	1.307
2D	0
3A	0.276
3B	0.089
3C	0.463
3D	0
4A	0.026
4B	0.605
4C	0.599
4D	0.642
5A	1.152
5B	1.359
5C	1.035
5D	0
6A	0.006
6B	1.201
6C	0.034
6D	1.227
7A	0.465
7B	0
7C	0.366
7D	0.77
8A	1.494
8B	1.427

8C	0.003
8D	1.413

Effect of transgenic expression of endoglucanase EglA on cell wall composition & in vitro digestibility analysis

Five individual seed from each of two lines, #263 & #266, not expressing or expressing EglA (pNOV4803) respectively were grown in the greenhouse. Protein extracts made from small leaf samples from immature plants were used to verify that transgenic endoglucanase activity was present in #266 plants but not #263 plants. At full plant maturity, ~30 days after pollination, the whole above ground plant was harvested, roughly chopped, and oven dried for 72 hours. Each sample was divided into 2 duplicate samples (labelled A & B respectively), and subjected to in vitro digestibility analysis using strained rumen fluid using common procedures (Forage fiber analysis apparatus, reagents, procedures, and some applications, by H. K. Goering and P. J. Van Soest, Goering, H. Keith 1941 (Washington, D.C.) : Agricultural Research Service, U.S. Dept. of Agriculture, 1970. iv, 20 p. : ill. -- Agriculture handbook ; no. 379 ), except that material was treated by a pre-incubation at either 40°C or 90°C prior to in vitro digestibility analysis. In vitro digestibility analysis was performed as follows:

Samples were chopped to about 1mm with a wiley mill, and then sub-divided into 16 weighed aliquots for analysis. Material was suspended in buffer and incubated at either 40°C or 90°C for 2 hours, then cooled overnight. Micronutrients, trypticase & casein & sodium sulfite were added, followed by strained rumen fluid, and incubated for 30 hours at 37°C. Analyses of neutral detergent fiber (NDF), acid detergent fiber (ADF) and acid detergent lignin (AD-L) were performed using standard gravimetric methods (Van Soest & Wine, Use of Detergents in the Analysis of fibrous Feeds. IV. Determination of plant cell-wall constituents. P.J. Van Soest & R.H. Wine. (1967). Journal of The AOAC, 50: 50-55; see also Methods for dietary fiber, neutral detergent fiber and nonstarch polysaccharides in relation to animal nutrition (1991). P.J. Van Soest, J.B. Robertson & B.A. Lewis. J. Dairy Science, 74: 3583-3597.).

Data show that transgenic plants expressing EglA (#266) contain more NDF than control plants (#233), whilst ADF & lignin are relatively unchanged. The NDF fraction of transgenic

plants is more readily digested than that of non-transgenic plants, and this is due to an increase in the digestibility of cellulose (NDF – ADF – AD-L), consistent with “self-digestion” of the cell-wall cellulose by the transgenically expressed endoglucanase enzyme.

#### **Example 45**

##### Expression of the thermophilic beta-glucanase/mannanase (6GP1) in corn

Transgenic seed for pNOV4820 and pNOV4823 were analysed for 6GP1 beta glucanase activity using the azo-barley glucan method (megazyme). Enzymatic assays conducted at 50 degrees C indicate that the transgenic seed have thermophilic 6GP1 beta-glucanase activity while no activity was detected in non-transgenic seed (positive signal represents background noise associated with this assay).

Table 11 shows activity of the thermophilic beta-glucanase/mannanase 6GP1 in transgenic corn seed. Transgenic seed for pNOV4820 (events 1-6) and pNOV4823 (events 7-9) were analysed for 6GP1 beta-glucanase activity using the azo-barley glucan method (megazyme). Enzymatic assays were conducted at 50 degrees C and the results indicate that the transgenic seed have thermophilic 6GP1 beta-glucanase activity while no activity is detected in non-transgenic seed.

Table 11

<b>Seed</b>	<b>Abs 590</b>
Wild Type	0
1	0.21
2	0.31
3	0.36
4	0.23
5	0.16
6	0.14
7	0.52
8	0.54
9	0.49

#### **Example 46**

Expression of the mesophilic barley AmyI amylase in corn

A variety of constructs were generated for the expression of the barley AmyI alpha-amylase in corn. The maize  $\gamma$ -zein and globulin promoters were used to express the amylase specifically in the endosperm or embryo, respectively. In addition, the maize  $\gamma$ -zein signal sequence and a synthetic ER retention signal were used to regulate the subcellular localization of the amylase protein. All 5 constructs (pNOV4867, pNOV4879, pNOV4897, pNOV4895, pNOV4901) yielded transgenic plants with alpha-amylase activity detected in the seed. Table 12 shows the activity in individual seed for 5 independent, segregating events (constructs pNOV4879 and pNOV4897). All of the constructs produced some transgenic events with a shrivelled seed phenotype indicating that synthesis of the barley AmyI amylase could effect starch formation, accumulation, or breakdown.

Table 12 shows activity of the barley AmyI alpha-amylase in individual corn seed (constructs pNOV4879 and pNOV4897). Individual, segregating seed for constructs pNOV4879 (seed samples 1 and 2) and pNOV4897 (seed samples 3-5) were analysed for alpha-amylase activity as described previously.

Table 12

Seed	U/g corn flour
1A	19.29
1B	1.49
1C	18.36
1D	1.15
1E	1.62
1F	14.99
1G	1.88
1H	1.83
2A	2.05
2B	36.79

2C	30.11
2D	2.25
2E	32.37
2F	1.92
2G	20.24
2H	35.76
3A	22.99
3B	1.72
3C	25.38
3D	18.41
3E	28.51
3F	2.11
3G	16.67
3H	1.89
4A	1.57
4B	36.14
4C	23.35
4D	1.70
4E	1.94
4F	14.38
4G	2.09
4H	1.83
5A	11.64
5B	18.20
5C	1.87
5D	2.07
5E	1.71
5F	1.92
5G	12.94
5H	15.25

**Example 47**



### Preparation of Xylanase Constructs

Table 13 lists 9 binary vectors that each contain a unique xylanase expression cassette. The xylanase expression cassettes include a promoter, a synthetic xylanase gene (coding sequence), an intron (PEPC, inverted), and a terminator (35S).

Two synthetic maize-optimized endo-xylanase genes were cloned into binary vector pNOV2117. These two xylanase genes were designated BD7436 (SEQ ID NO: 61) and BD6002A (SEQ ID NO:63). Additional binary vectors containing a third maize-optimized sequence, BD6002B (SEQ ID NO:65) can be made.

Two promoters were used: the maize glutelin-2 promoter (27-kD gamma-zein promoter (SEQ ID NO: 12 ) and the rice glutelin-1 (Osgt1) promoter (SEQ ID NO: 67). The first 6 vectors listed in Table 1 have been used to generate transgenic plants. The last 3 vectors can also be made and used to generate transgenic plants.

Vector 11560 and 11562 encode the polypeptide shown in SEQ ID NO: 62 (BD7436). Constructs 11559 and 11561 encode a polypeptide consisting of SEQ ID NO: 17 fused to the N-terminus of SEQ ID NO: 62. SEQ ID NO: 17 is the 19 amino acid signal sequence from the 27-kD gamma-zein protein.

Vector 12175 encodes the polypeptide shown in SEQ ID NO: 64(BD6002A). Vector 12174 encodes a fusion protein consisting of the gamma-zein signal sequence (SEQ ID NO: 17) fused to the N-terminus of SEQ ID NO: 64.

Vectors pWIN062 and pWIN064 encode the polypeptide shown in SEQ ID NO: 66(BD6002B). Vector pWIN058 encodes a fusion protein consisting of the chloroplast transit peptide of maize waxy protein (SEQ ID NO:68) fused to the N-terminus of SEQ ID NO: 66 .

Table 13 Xylanase binary vectors

Vector	Promoter	Signal Sequence Source	Xylanase Gene
11559	27kD Gamma-zein	27kD Gamma-zein	BD7436
11560	27kD Gamma-zein	None	BD7436
11561	OsGt1	27kD Gamma-zein	BD7436
11562	OsGt1	None	BD7436
12174	27kD Gamma-zein	27kD Gamma-zein	BD6002A
12175	27kD Gamma-zein	None	BD6002A
PWIN058	27kD Gamma-zein	Maize waxy protein	BD6002B
PWIN062	OsGt1	None	BD6002B
PWIN064	27kD Gamma-zein	None	BD6002B

All constructs include an expression cassette for PMI, to allow positive selection of regenerated transgenic tissue on mannose-containing media.

#### Example 48

##### Xylanase Activity Assay Results

The data shown in Tables 14 and 15 demonstrate that xylanase activity accumulates in T1 generation seed harvested from regenerated (T0) maize plants stably transformed with binary vectors containing xylanase genes BD7436 (SEQ ID NO: 61 in Example 47) and BD6002A (SEQ ID NO:63 in Example 47). Using an Azo-WAXY assay (Megazyme), activity was detected in extracts from both pooled (segregating) transgenic seed and single transgenic seed.

T1 seed were pulverized and soluble proteins were extracted from flour samples using citrate-phosphate buffer (pH 5.4). Flour suspensions were stirred at room temperature for 60 minutes, and insoluble material was removed by centrifugation. The xylanase activity of the supernatant fraction was measured using the Azo-WAXY assay (McCleary, B.V. "Problems in the measurement of beta-xylanase, beta-glucanase and alpha-amylase in feed enzymes and animal feeds". In proceedings of Second European Symposium on Feed Enzymes" (W.van Hartingsveldt, M. Hessing, J.P. van der Jugt, and W.A.C Somers Eds.), Noordwijkerhout, Netherlands, 25-27 October, 1995). Extracts and substrate were pre-incubated at 37°C. To 1 volume of 1X extract supernatant, 1 volume of substrate (1% Azo-Wheat Arabinoxylan S-AWXP) was added and then incubated at 37°C for 5 minutes. Xylanase activity in the corn

flour extract depolymerizes the Azo-Wheat Arabinoxylan by an endo-mechanism and produces low molecular weight dyed fragments in the form of xylo-oligomers. After the 5 minute incubation, the reaction was terminated by the addition of 5 volumes of 95% EtOH. Addition of alcohol causes the non-depolymerized dyed substrate to precipitate so that only the lower molecular weight xylo-oligomers remain in solution. Insoluble material was removed by centrifugation. The absorbance of the supernatant fraction was measured at 590nm, and the units of xylanase per gram of flour were determined by comparison to the absorbance values from identical assays using a xylanase standard of known activity. The activity of this standard was determined by a BCA assay. The enzyme activity of the standard was determined using wheat arabinoxylan as substrate and measuring the release of reducing ends by reaction of the reducing ends with 2,2'-bicinchoninic acid (BCA). The substrate was prepared as a 1.4% w/w solution of wheat arabinoxylan (Megazyme P-WAXYM) in 100 mM sodium acetate buffer pH5.30 containing 0.02% sodium azide. The BCA reagent was prepared by combining 50 parts reagent A with 1 part reagent B (reagents A and B were from Pierce, product numbers 23223 and 23224, respectively). These reagents were combined no more than four hours before use. The assay was performed by combining 200 microliters of substrate to 80 microliters of enzyme sample. After incubation at the desired temperature for the desired length of time, 2.80 milliliters of BCA reagent was added. The contents were mixed and placed at 80°C for 30-45 minutes. The contents were allowed to cool and then transferred to cuvettes and the absorbance at 560nm was measured relative to known concentrations of xylose. The choice of enzyme dilution, incubation time, and incubation temperature could be varied by one skilled in the art.

The experimental results shown in Table 14 demonstrate the presence of recombinant xylanase activity in flour prepared from T1 generation corn seed. Seed from 12 T0 plants (derived from independent T-DNA integration events) were analyzed. The 12 transgenic events were derived from 6 different vectors as indicated (refer to Table 13 in Example 47 for description of vectors). Extracts of non-transgenic (negative control) corn flour do not contain measurable xylanase activity (see Table 15). The xylanase activity in these 12 samples ranged from 10-87 units/gram of flour.

Table 14. Analysis of pooled T1 seed.

Vector	Sample	Xylanase Units / Gram of Flour
11559	MD9L013800	63
11559	MD9L012428	58
11560	MD9L011296	33
11560	MD9L011322	21
11561	MD9L012413	87
11561	MD9L012443	83
11562	MD9L012890	13
11562	MD9L013788	12
12174	MD9L022080	16
12174	MD9L022195	10
12175	MD9L022061	74
12175	MD9L022134	69

The results in Table 15 demonstrate the presence of xylanase activity in corn flour derived from single kernels. T1 seed from two T0 plants containing vectors 11561 and 11559 were analyzed. These vectors are described in Example 47. Eight seed from each of the two plants were pulverized and flour samples from each seed were extracted. The table shows results of single assays of each extract. No xylanase activity was found in assays of extracts of seeds 1, 5, and 8 for both transgenic events. These seed represent null segregants. Seed 2, 3, 4, 6, and 7 for both transgenic events accumulated measurable xylanase activity attributable to expression of the recombinant BD7436 gene. All 10 seed that tested positive for xylanase activity (>10 unit/gram flour) had an obvious shriveled or shrunken appearance. By contrast the 6 seed that tested negative for xylanase activity ( $\leq 1$  unit/gram flour) had a normal appearance. This result suggests that the recombinant xylanase depolymerized endogenous (arabino)xylan substrate during seed development and/or maturation.

Table 15. Analysis of single T1 seed.

Vector 11561	Vector 11559
--------------	--------------

Seed Number	Xylanase Units / Gram of Flour	Seed Number	Xylanase Units / Gram of Flour
1	0	1	1
2	45	2	52
3	38	3	21
4	40	4	13
5	0	5	0
6	40	6	28
7	32	7	23
8	0	8	0

### Example 49

#### Enhanced starch recovery from corn seed using enzymes

Corn wet-milling includes the steps of steeping the corn kernel, grinding the corn kernel, and separating the components of the kernel. A bench top assay (the Cracked Corn Assay) was developed to mimic the corn wet-milling process

The "Cracked Corn Assay" was used for identifying enzymes that enhance starch yield from maize seed resulting in an improved efficiency of the corn wet milling process. Enzyme delivery was either by exogenous addition, transgenic corn seed, or a combination of both. In addition to the use of enzymes to facilitate separation of the corn components, elimination of SO<sub>2</sub> from the process is also shown.

#### Cracked Corn Assay.

One gram of seed was steeped overnight in 4000, 2000, 1000, 500, 400, 40, or 0 ppm SO<sub>2</sub> at 50 degrees C or 37 degrees C. Seeds were cut in half and the germ removed. Each half seed was cut in half again. Steep water from each steeped seed sample was retained and diluted to a final concentrations ranging from 400 ppm to 0 ppm SO<sub>2</sub>. Two milliliters of the steep water with or without enzymes was added to the de-germed seeds and the samples placed at 50

degrees C or 37degrees C for 2-3 hours. Each enzyme was added at 10 units per sample. All samples were vortexed approximately every 15 minutes. After 2-3 hours the samples were filtered through mira cloth into a 50ml centrifuge tube. The seeds were washed with 2 ml of water and the sample pooled with the first supernatant. The samples were centrifuged for 15 minutes at 3000 rpm. Following centrifugation, the supernatant was poured off and the pellet placed at 37 degrees C to dry. All pellet weights were recorded. Starch and protein determinations were also carried out on samples for determining the starch:protein ratios released during the treatments (data not shown).

Analysis of T1 and T2 seed from maize plants expressing 6GP1 endoglucanase in Cracked corn Assay

Transgenic corn (pNOV4819 and pNOV4823) containing a thermostable endoglucanase performed well when analyzed in the Cracked Corn Assay. Recovery of starch from the pNOV4819 line was found to be 2 fold higher in seeds expressing the endoglucanase when steeped in 2000 ppm SO<sub>2</sub>. Addition of a protease and cellobiohydrolase to the endoglucanase seed increased the starch recovery approximately 7 fold over control seeds. See Table 16.

Table 16. Crack Corn Assay results for cytosolic expressed Endoglucanase (pNOV4820). Control line, A188/HiII PNOV4819 lines, 42C6A-1-21 and 27.

Maize Line	Treatment	Starch Pellet Wt. (mg)
A188/HiII Control	No Enzyme	28.4
A188/HiII Control	Bromelain/C8546 10U	109.3
42C6A-1-21	No Enzyme	52.6
42C6A-1-21	Bromelain/C8546 10U	170.4
42C6A-1-27	No Enzyme	60.5
42C6A-1-27	Bromelain/C8546 10U	207.5

Similar results were seen in transgenic seed containing endoglucanase targeted to the ER of the

endosperm (pNOV4823), again resulting in a 2 –7 fold increase in starch recovery when compared to control seed. See Table 17.

Table 17. Crack Corn Assay results for ER expressed endoglucanase (pNOV4823). Control line, A188/Hill; PNOV4823 line, 101D11A-1-28.

Line	Treatment	Starch Pellet Wt (mg)	Starch Pellet Wt (mg)	Mean Wt.
A188/Hill	No Enzyme	22.5	19.1	20.8
101D11A-1-28	No Enzyme	41.2	32	36.6
A188/Hill	10U Bromelian/C8546	78.6	73.8	76.2
101D11A-1-28	10U Bromelian/C8546	169.8	132.6	151.2

These results confirm that expression of an endoglucanase enhances the separation of starch and protein components of the corn seed. Further more it could be shown that reduction or removal of SO<sub>2</sub> during the steeping process resulted in starch recovery that was comparable to or better than normally steeped control seeds. See Table 18. Removal of high levels of SO<sub>2</sub> from the wet-milling process can provide value-added benefits.

Table 18. Comparison of various concentrations of SO<sub>2</sub> on starch recovery from transgenic 6GPI seed.

Line	Treatment	Starch Pellet Wt (mg)
A188 Control	2000 ppm SO <sub>2</sub>	18.5
JHAF Control	2000 ppm SO <sub>2</sub>	29.1
42C (pNOV4820)	2000 ppm SO <sub>2</sub>	29.5
101C (pNOV4823)	2000 ppm SO <sub>2</sub>	73.1
101D (pNOV4823)	2000 ppm SO <sub>2</sub>	42.5
136A (pNOV4825)	2000 ppm SO <sub>2</sub>	36.6

137A (pNOV4825)	2000 ppm SO <sub>2</sub>	38.6
42C (pNOV4820)	400 ppm SO <sub>2</sub>	18.5
101C (pNOV4823)	400 ppm SO <sub>2</sub>	20.4
101D (pNOV4823)	400 ppm SO <sub>2</sub>	39.7
136A (pNOV4825)	400 ppm SO <sub>2</sub>	26
137A (pNOV4825)	400 ppm SO <sub>2</sub>	26.9
42C (pNOV4820)	0 ppm SO <sub>2</sub>	21.9
101C (pNOV4823)	0 ppm SO <sub>2</sub>	32.5
101D (pNOV4823)	0 ppm SO <sub>2</sub>	39
136A (pNOV4825)	0 ppm SO <sub>2</sub>	17.8
137A (pNOV4825)	0 ppm SO <sub>2</sub>	29.2

### Example 50

#### Construction of transformation vectors for maize optimized bromelain

Expression cassettes were constructed to express the maize optimized bromelain in maize endosperm with various targeting signals as follows:

pSYN11000 (SEQ ID NO. 73 ) comprises the bromelain signal sequence (MAWKVQVVFLFLFLCVMWASPSAASA) (SEQ ID NO: 72) and synthetic bromelain sequence fused with a C-terminal addition of the sequence VF AE A I A A N S T L V A E for targeting to and retention in the PVS (Vitale and Raikhel Trends in Plant Science Vol 4 no.4 pg 149-155). The fusion was cloned behind the maize gamma zein promoter for expression specifically in the endosperm.

pSYN11587 (SEQ ID NO:75) comprises the bromelain N-terminal signal sequence (MAWKVQVVFLFLFLCVMWASPSAASA) and synthetic bromelain sequence with a C-terminal addition of the sequence SEKDEL for targeting to and retention in the endoplasmic reticulum (ER) (Munro and Pelham, 1987). The fusion was cloned behind the maize gamma zein promoter for expression specifically in the endosperm.

pSYN11589 (SEQ ID NO. 74) comprises the bromelain signal sequence (MAWKVQVVFLFLFLCVMWASPSAASA) (SEQ ID NO: 72) fused to the lytic vacuolar



targeting sequence SSSSFADSNPIRVTDRRAAST (Neuhaus and Rogers Plant Molecular Biology 38:127-144, 1998) and synthetic bromelain for targeting to the lytic vacuole. The fusion was cloned behind the maize gamma zein promoter for expression specifically in the endosperm.

pSYN12169 (SEQ ID NO: 76) comprises the maize  $\gamma$ -zein N-terminal signal sequence (MRVLLVALALLALAASATS)(SEQ ID NO:17) fused to the synthetic bromelain for targeting to the endoplasmic reticulum and secretion into the apoplast (Torrent et al. 1997). The fusion was cloned behind the maize gamma zein promoter for expression specifically in the endosperm.

pSYN12575 (SEQ ID NO:77) comprises the waxy amyloplast targeting peptide (Klosgen et al., 1986) fused to the synthetic bromelain for targeting to the amyloplast. The fusion was cloned behind the gamma zein promoter for expression specifically in the endosperm.

pSM270 ( SEQ ID NO.78 ) comprises the bromelain N-terminal signal sequence fused to the lytic vacuolar targeting sequence SSSSFADSNPIRVTDRRAAST (Neuhaus and Rogers Plant Molecular Biology 38:127-144, 1998) and synthetic bromelain for targeting to the lytic vacuole. The fusion was cloned behind the aleurone specific promoter P19 (US Patent 6392123) for expression specifically in the aleurone.

### Example 51

#### Expression of bromelain in corn

Seeds from T1 transgenic lines transformed with vectors containing the synthetic bromelain gene with targeting sequences for expression in various subcellular location of the seed were analyzed for protease activity. Corn-flour was made by grinding seeds, for 30 sec., in the Kleco grinder. The enzyme was extracted from 100 mg of flour with 1 ml of 50 mM NaOAc pH4.8 or 50 mM Tris pH 7.0 buffer containing 1mM EDTA and 5 mM DTT. Samples were vortexed, then placed at 4C with continuous shaking for 30 min. Extracts from each transgenic line was assayed using resorufin labeled casein (Roche, Cat. No. 1 080 733) as outlined in the product brochure. Flour from T2 seeds were assayed using a bromelain specific assay as outlined in Methods in Enzymology Vol. 244: Pg 557-558 with the following modifications. 100mg of corn seed flour was extracted with 1ml of 50mMNa<sub>2</sub>HPO<sub>4</sub>/50mM NaH<sub>2</sub>PO<sub>4</sub>, pH 7.0, 1

mM EDTA +/- 1 $\mu$ M leupeptin for 15 min at 4°C. Extracts were centrifuged for 5 min at 14,000 rpm at 4°C. Extracts were done in duplicates. Flour from T2 Transgenic lines was assayed for bromelain activity using Z-Arg-Arg-NHMec (Sigma) as a substrate. Four aliquots of 100 $\mu$ l /corn seed extracts were added to 96 well flat bottom plates (Corning) containing 50 $\mu$ l 100mMNa<sub>2</sub>HPO<sub>4</sub>/100mM NaH<sub>2</sub>PO<sub>4</sub>, pH 7.0, 2mM EDTA, 8mM DTT/well. The reaction was started by the addition of 50 $\mu$ l of 20 $\mu$ M Z-Arg-Arg-NHMec. The reaction rate was monitor using a SpectraFluorPlus(Tecan) fitted with a 360nm excitation and 465nm emission filters at 40°C at 2.5min intervals.

Table 19 shows the analysis of seed from different T1 bromelain events. Bromelain expression was found to be 2-7 fold higher than the A188 and JHAF control lines. T1 transgenic lines were replanted and T2 seeds obtained. Analysis of T2 seeds showed expression of bromelain. Figure 21 shows bromelain activity assay using Z-Arg-Arg-NHMec in T2 seed for ER targeted (11587) and lytic vacuolar targeted (11589) bromelain.

*Analysis of T2 seed from maize plants expressing Bromelain*

Seed from T2 transgenic bromelain line, 11587-2 was analyzed in the Cracked Corn assay for enhanced starch recovery. Previous experiments using exogenously added bromelain showed an increased starch recovery when tested alone and in combination with other enzymes, particularly cellulases. The T2 seed from line 11587-2 showed a 1.3 fold increase in starch recovered over control seed when steeped at 37C/2000 ppm SO<sub>2</sub> overnight. More importantly, there was the 2 fold increase in starch from the T2 bromelain line, 11587-2 when a cellulase (C8546) was added when seeds were steeped at 37C/2000 ppm SO<sub>2</sub>.

The transgenic line showed a similar trend in increased starch over control seed when seeds were steeped at 37C/400 ppm SO<sub>2</sub>. A 1.6 fold increase starch recovered over control was

seen in the transgenic seed and a 2.1 fold increase of starch with addition of a cellulase (C8546).

See Table 20.

These results are significant in showing that it is possible to reduced temperature and SO<sub>2</sub> levels while also enhancing the starch recovery during the wet-milling process when transgenic seed expressing a bromelain is used.

Table 19

Summary of Grain Specific Expression of Bromelain in T1 corn.

Line Number	Targeting	Construct	"Specific Activity" ng Bromelain/protein
11000-1	Vacuolar	GZP/probromelain/barleyPVS	252
11000-2	Vacuolar	GZP/probromelain/barleyPVS	277
11000-3	Vacuolar	GZP/probromelain/barleyPVS	284
11587-1	ER	GZP/probromelain/KDEL	174
11587-1	ER	GZP/probromelain/KDEL	153
11589-1	Lytic Vacuolar	GZP/aleurainSS/probromelain	547
11589-2	Lytic Vacuolar	GZP/aleurainSS/probromelain	223
		A188 Control	56
		JHAF Control	75

Table 20 Cracked Corn Assay results for T2 Bromelain seed

Steep Conditions	Line	Starch Pellet Wt. (mg)
2000 ppm SO <sub>2</sub>	A188	41.3
2000 ppm SO <sub>2</sub>	A188/C8546 (10 units)	44
2000 ppm SO <sub>2</sub>	11587-2	57.4
2000 ppm SO <sub>2</sub>	11587-2/C8546 (10 units)	94.6
400 ppm	A188	30.7
400 ppm	A188/C8546 (10 units)	35.8
400 ppm	11587-2	50.5
400 ppm	11587-2/C8546 (10 units)	86.6

**Example 52****Construction of transformation vectors for maize optimized ferulic acid esterase.**

Expression cassettes were constructed to express the maize optimize ferulic acid esterase in maize endosperm with or without various targeting signals as follows:

Plasmid 13036 (SEQ ID NO: 101) comprises the maize optimize ferulic acid esterase (FAE) sequence (SEQ ID NO: 99). The sequence was cloned behind the maize gamma zein promoter without any targeting sequences for expression specifically in the cytosol of the endosperm.

Plasmid 13038 (SEQ ID NO: 103) comprises the maize  $\gamma$ -zein N-terminal signal sequence (MRVLLVALALLALAASATS)(SEQ ID NO:17) fused to the synthetic FAE for targeting to the endoplasmic reticulum and secretion into the apoplast (Torrent et al. 1997). The fusion was cloned behind the maize gamma zein promoter for expression specifically in the endosperm.

Plasmid 13039 (SEQ ID NO: 105) comprises the waxy amyloplast targeting peptide (MLAALATSQLVATRAGLGVPDASTFRRGAAQGLRGARASAAAD TL SMRTSARAAPRHQHQARRGARFPSLVVCASAGA) (Klosgen et al., 1986) fused to the synthetic FAE for targeting to the amyloplast. The fusion was cloned behind the gamma zein promoter for expression specifically in the endosperm.

Plasmid 13347 (SEQ ID NO: 107) comprises the maize  $\gamma$ -zein N-terminal signal sequence (MRVLLVALALLALAASATS)(SEQ ID NO:17) fused to the synthetic FAE sequence with a C-terminal addition of the sequence SEKDEL for targeting to and retention in the endoplasmic reticulum (ER) (Munro and Pelham, 1987). The fusion was cloned behind the maize gamma zein promoter for expression specifically in the endosperm.

All expression cassettes were moved into a binary vector pNOV2117 for transformation into maize via *Agrobacterium* infection. The binary vector contained the phosphomannose isomerase (PMI) gene which allows for selection of transgenic cells with mannose. Transformed maize plants were either self-pollinated or outcrossed and seed was collected for analysis.

Combinations of the enzymes can be produced either by crossing plants expressing the individual enzymes or by cloning several expression cassettes into the same binary vector to enable cotransformation.

Synthetic Ferulic Acid Esterase Sequence (SEQ ID NO: 99)

atggccgctccctcccgaccatgccgcgtccggtacgaccaggtgcgcaacggcgtgccgcggccagggtgggaacatctctacttctccaccgccacaa  
ctccacccgccgcggccggtgtacctcccgccgggtactccaaggacaagaagtactccgtgctctacctctccacggcatcgccggtccggagaaacgactggt  
cgaggcgccgcccgcgccaacgtgatccgacaaacctatcgccgagggaagatcaagccgtcatcatcgtgaccccgaaacaccaacggccggccggccg  
gcatcccgacgggtacgagaacttccaaggacctctcaactccctatcccgatcatcgagtcacactaccgtgtacaccgaccgcgagcaccgcgccatgc  
cggcctctctatggcgccggccagtcctcaacatcgccctcacaacctcgacaagttcgctacatcgcccgatctccgcccgaacacclaccggaacga  
cgccctctcccggaaggcgcaaggccgcccgcgagaagctcaagctctctctatcgctcgccgaccaacgactccctcatcggttcggccagcgcggtgcacg  
agtactcggtggccaacaacatcaaccacgtgtactggtctatccaggcgccggccacgacttcaacgtgtggaagccggccctctggaacttctccagatggccg  
acgaggccggcctaccgcgacggcaacacccgggtgccgaccccgctcccgaagccggccaacacccgcatcgaggccgaggactacgacggcatcaactcc  
tctctcatcgagatcatcggtgcccggcgaggcgccgcccgcgcatcggtacatcacctccggcgactacclcggtacaagtcacgacttcggcaacggccgc  
acctctcaagggcaagggtggccaacgccaacacctccaacatcgagcttcgctcaacggcccgaacggcacccctcatcgccaccctctccgtgaagtccaccggc  
gactggaacacctacgaggagcagacctgctccatctcaaggtgaccggcatcaacgaccttacctctgttcaaggcccggtgaacatcgactggttcaacttcg  
cggtgtag

Synthetic Ferulic Acid Esterase Amino Acid Sequence (SEQ ID NO: 100)

maasltmppsdydqvrmgvprrgvvnisyfstamstrparvylppgyskdkkysvlyllhgiggssendwfeegggranviadnliaegkikplliivtpntmaagp  
giadgyenftkdllnslipyiesnysvytdrehraiaglsmgggqsfngltmldkfayigpisaapntypnerlfpdggaareklklfiacgtmdsligfgqrvheyc  
vanninhvywliqggghdfnvkpglwnflqmadeagltrdntpvtpspkpantrieaedydginssieiiigvppeggrgigvitsgdylyvksidfngat  
sfkakvanantsnielrlngpntligtllsvkstgdwntyeegtcisiskvtgindlylvfkpgpnidwftfgv\*

13036 Sequence (SED ID NO: 101)

atggccgctccctcccgaccatgccgcgtccggtacgaccaggtgcgcaacggcgtgccgcggccagggtgggaacatctctacttctccaccgccacaa  
ctccacccgccgcggccggtgtacctcccgccgggtactccaaggacaagaagtactccgtgctctacctctccacggcatcgccggtccggagaaacgactggt  
cgaggcgccgcccgcgccaacgtgatccgacaaacctatcgccgagggaagatcaagccgtcatcatcgtgaccccgaaacaccaacggccggccggccg  
gcatcccgacgggtacgagaacttccaaggacctctcaactccctatcccggtacatcgagtcacactactccgtgtacaccgaccgcgagcaccgcgccatgc  
cggcctctctatggcgccggcgccagtcctcaacatcgccctcacaacctcgacaagttcgctacatcgcccgatctccgcccgaacacctaccgaacga  
cgccctctccggagcgccgcaaggccggccgcgagaagctcaagctctctctatcgccctggcaccacgactccctcatcggttcggccagcgcggtgcacg  
agtactcggtggccaacaacatcaaccacgtgtactggtctatccaggcgccggccacgacttcaacgtgtggaagccggccctctggaacttctccagatggccg  
acgaggccggcctaccgcgacggcaacacccgggtgccgaccccgctcccgaagccggccaacacccgcatcgaggccgaggactacgacggcatcaactcc  
tctccatcgagatcatcggtgcccggcgaggcgccgcccgcgcatcggtacatcacctccggcgactacclcggtacaagtcacgacttcggcaacggccgc  
acctctcaagggcaagggtggccaacgccaacacctccaacatcgagcttcgctcaacggcccgaacggcacccctcatcgccaccctctccgtgaagtccaccggc  
gactggaacacctacgaggagcagacctgctccatctcaaggtgaccggcatcaacgaccttacctctgttcaaggcccggtgaacatcgactggttcaacttcg  
cggtgtag

13036 AA Sequence (SED ID NO: 102)

maasltmppsdydqvrmgvprrgvvnisyfstamstrparvylppgyskdkkysvlyllhgiggssendwfeegggranviadnliaegkikplliivtpntmaagp  
giadgyenftkdllnslipyiesnysvytdrehraiaglsmgggqsfngltmldkfayigpisaapntypnerlfpdggaareklklfiacgtmdsligfgqrvheyc  
vanninhvywliqggghdfnvkpglwnflqmadeagltrdntpvtpspkpantrieaedydginssieiiigvppeggrgigvitsgdylyvksidfngat  
sfkakvanantsnielrlngpntligtllsvkstgdwntyeegtcisiskvtgindlylvfkpgpnidwftfgv\*

13038 Sequence (SEQ ID NO: 103)

atgagggtgtgctgttgccctcgctctctggtctcgctcgagcgccacccatggccgctccctcccgaccatgcccgccgtccggtacgaccaggtgccga  
acggcgtgccgcggccagggtggtgaacatctctacttctccaccgccaccaactccacccgcccggcccgcgtgtacctcccgccggctactccaaggacaag

aagtlactcgtgcttacctctccacggcatcggcggtccgagaacgactggttcgagggcgcgcccgcccaacgtatcggcgacaacctatcgccgaggg  
caagatcaagcgcctatcatcgtgacccgaacacacacggcgccggccggcgatcggcgacggctacgagaacttaccaggacctctcaactccctcatccc  
gtacatcgagtcacactactcgtgtacaccgaccgagcaccggccatcggcggtctctatggcgggcgccagtccttcaacatcgccctcaccaccccgac  
aagttcgctatcatcgcccgatcctccgccccgaacacctaccggaacgagcgctcttccggacggcggaagggcgcccgagaaagctcaagctcctt  
catcgctcggcgacccaacgactccctcatcggctcggccagcggtgcacgagtagctggtggccaacaacatcaaccagtgtagtggctcatccaggcgcg  
ccacgactcaacgtgtggaagccggcgctctggaacttctccagatggccgacgagggcgccctaccggcgacggcaacaccccggtgccgacccgctcccg  
aagccggccaacacccgcacgagggcgaggactacgacggcatcaactctctccatcagatcatcggtggcgccggagggcgcccgccgcatcgctac  
atcactcggcgactacccgtgtacaaagtcacgacttcggcaacggcgccacctcttcaaggccaaggtggccaacggcaacacctccaacatcgagcttcgcc  
tcaacggcccggaacggcaccctcatcgccacctctcgtggaagtcaccggcgactggaaacacctacgaggagcagacctgtctcatctccaaggtgacccgcatc  
aacgacctctactctgtgttcaaggcccggtgaacatcgactggttacccttcggcggttag

#### 13038 AA Sequence (SEQ ID NO: 104)

mrllvalallalaasatsmaaslpmpspgydqvrngvprgqvvnisyfstatmstrparvylppgyskdkkysvlyllhgiggsewdfeggrnviadnliae  
gkikpliivtpntnaagpiadgyenftkdlinslipiesnysvtdrehraiaaglsmgggqsfniqlmldkfayigpisaapntypnerlfpdggkaareklklfia  
cgtndsligfgqrvheycvanninhvwyliqggghdfnvwkpglwnflqmadeagltrdgnptvptpspkpantrieaedydginssieiiyvppeggrgigvi  
tsedylyvksidfngatsfkakvanantsnielrlnpgnngltglsvkstgdwntyeetcsiskvtgindlylvfkpgpvnidwftfgv\*

#### 13039 Sequence (SEQ ID NO: 105)

atgctggcggtctgcccagctgcagctcgtcgcaacggcgccggcctggcggtccggagcggtccacgttccggcgggcgccggcgagggcctgagggg  
ggccggcggtcggcgggcgggacacgctcagcagcggaccagcgcgcgcgccggccagggcaccagcaccagcagggcgccggcgggcgaggttcc  
cgctcgtcgtcgtgtgctggcagcgccggcgccatggcggtcctctccgacctgcccggctcggctacgaccaggtggcgaacggcggtggcgcgccgaggt  
ggtgaacatctctactctccaccggcaccacacccggcgccggcggtgtagtctcccgccggctacccaaggacaagaagtagtccgtgctctacctctcc  
acggcatcgggcgctcggagaacgactggtcggagggcgccggcgcccaacgtgacggcgacaacctcatcgccgagggcaagatcaagccgctcatcatcgt  
gacccgaacccaacggcgccggcgccggcgatcgccgagcgctacgagaacttcaacaggacctctcaactccctcatccgtatcagtagtccaactactcctgt  
gtacacggcgccgagcaccggccatcgccggcctctctatggcgcgcgccagtccttcaacatcgccctcacaacctcgacaagttcgctcatcggccgpat  
ctccggcgcccggaacacctaccggaacgagcgctcttccggagcgcggaaggcgcccgagaaagctcaagctccttctcatcgctcggcgccaacgact  
ccctcatcggtctggccagcggtgcacgagtagtctgctggccaacaacatcaaccagtgtagtggctcatcaggcgcgcgccacgacttcaacgtgtggagc  
cgggcctctggaactctctcagatggccgacgagggcgccctaccggcgacggcaacacccgggtggcgaccccgctcccggaagcgccgaacacccgcatcg  
aggcgaggactacgagggcatcaactcctctccatcgagatcatcggtggcgccggagggcgccggcgatcggtacatcactcggcgactacctcgtg  
tacaagtcacgacttcggcaacggcgccacctcttcaaggccaaggtggccaacggcaacacctccaacatcgaggttcgctcaacggcgccgaacggcacctc  
atcgccacctctcctggaagttccaccggcgacttgaacacctagaggagcagacctgctccatctccaagggtgacggcgatcaacgaccttcaactcgtgttcaagg  
gcccgtggaacatcgactgtttacaccttggcggttag

#### 13039 AA Sequence (SEQ ID NO: 106)

mlaalatsqlvatraglvpdastfrgaagqlrgarasaadilsmrtsaraaprhhqhqqarrgarfplsivcasagamaaslpmpspgydqvrngvprgqvvn  
syfstatmstrparvylppgyskdkkysvlyllhgiggsewdfeggrnviadnliaegkikpliivtpntnaagpiadgyenftkdlinslipiesnysvtdre  
hraiaaglsmgggqsfniqlmldkfayigpisaapntypnerlfpdggkaareklklfiacgtndsligfgqrvheycvanninhvwyliqggghdfnvwkpglw  
nflqmadeagltrdgnptvptpspkpantrieaedydginssieiiyvppeggrgigviitsedylyvksidfngatsfkakvanantsnielrlnpgnngltglsvk  
stgdwntyeetcsiskvtgindlylvfkpgpvnidwftfgv\*

#### 13347 Sequence (SEQ ID NO: 107)

atgaggggtgtgctgtgcccctgcctcctggctcgtcgcgagcgccacctccatggcgccctccctccgacctgcccggctcggctacgaccaggtgcga  
acggcggtccggcgcgccaggtgtggaacatctcctacttctccaccggcaccacacccggccggcggtgtacttccggcggtactccaaggacaag  
aagtagtaccgtgcttactctctccaccggcgatcgggcggtcggagaaagcagtaggttcgagggcgcgcccgcgccaacgtatcgccgacaacctcatcgccgaggg  
caagatcaagccgtcatcatcgtgacccgaacacacacggcgccggcgccggcgatcgccgacggctacgagaacttccaaggacctctcaactccctcatccc  
gtacatcgagttcaactactcgtgtacaccgaccggagcaccggccatcgccggcctctctatggcgggcgccagtccttcaacatcgccctcacaacctcgac  
aagttccttactatcgcccgatcctccgccccgaacacctaccggaacgagcgctcttccggagcgcggaagggcgcccgagaaagctcaagctcctt  
catcgctcggcgacccaacgactccctcatcggtctcggcagcggtgcacgagtagctggtggccaacaacatcaaccagtgtagtggctcatcaggcgcgccg  
ccacgacttcaacgtgtggaagccggcgctctggaacttctccagatggccgacgagggcgccctaccggcgacggcaacacccgggtgccgacccgltcccg  
aagccggccaacacccgcacgagggcgaggactacgacggcatcaactctctccatcgagatcatcggtggcgccggagggcgccggcgatcggttac  
atcactcggcgactacctcgtgtacaagtcacatcgacttcggcaacggcgccacctcttcaaggccaaggtggccaacggcaacacctccaacatcgagcttcgcc

tcaacgccccgaacggcaccctcatcggcaccctctccgtgaagtcaccggcgactggaacacctacgaggagcagacctgctccatctccaaggtagaccggcatc  
aacgacctctacctcgtgtcaaggccccgggtgaacatcgactggtcaccttcggcgtgtccgagaaggacgaactctag

13347 Sequence (SEQ ID NO: 108)

mrvtllvalallalaasatmaaslpmpsgydvrmgvprgqvvnisyfstatnstrparvylppgvskdkkysvlyllhgiggseendwfeggggranviadnliae  
gkikplliivtpntnaagpgiadgyenftkdllnslipyiesnysvtdrehraiaaglsmgggqsfniqlmldkfayigpisaapntypnerlfpdggkaareklklfia  
cgtmdsligfegrveyevanninhvywliqggghdfnvwkpglwnflqmadeagltrdgnptvptpspkpantriaedydginssieiigvppeggrigvi  
tsgdylvyksidfngatsfkakvanantsnielrlngpntglitlsvkstgdwntyeetcsiskvtgindlylvfkgpvnidwrtfgvsekdel\*

### Example 53

#### Hydrolytic degradation of corn fiber by ferulic acid esterase

Corn fiber is a major by-product of corn wet and dry milling. The fiber component is composed primarily of coarse fiber arising from the seed pericarp (hull) and aleurone, with a smaller fraction of fine fiber coming from the endosperm cell walls. Ferulic acid, a hydroxycinnamic acid, is found in high concentrations in the cell walls of cereal grains resulting in a cross linking of lignin, hemicellulose and cellulose components of the cell wall. Enzymatic degradation of ferulate cross-linking is an important step in the hydrolysis of corn fiber and may result in the accessibility of further enzymatic degradation by other hydrolytic enzymes.

#### Ferulic Acid Esterase Activity Assay

Ferulic acid esterase, FAE-1, (maize optimised synthetic gene from *C. thermocellum*) was expressed in *E. coli*. Cells were harvested and stored at  $-80^{\circ}\text{C}$  overnight. Harvested bacteria was suspended in 50mM Tris buffer pH7.5. Lysozyme was added to a final concentration of 200  $\mu\text{g/mL}$  and the sample incubated 10 minutes at room temperature with gently shaking. The sample was centrifuged at  $4^{\circ}\text{C}$  for 15 minutes at 4000 rpm. Following centrifugation, the supernatant was transferred to a 50 mL conical tube, and placed in 70 degree Celsius water bath for 30 minutes. The sample was then centrifuged for 15 minutes at 4000 rpm and the cleared supernatant transferred to a conical tube (Blum et al. J Bacteriology, Mar 2000, pg 1346-1351.)

The recombinant FAE-1 was tested for activity using 4-methylumbelliferyl ferulate as described in Mastihubova et al (2002) Analytical Biochemistry 309 96-101. Recombinant protein FAE-1 (104-3) was diluted 10, 100, and 1000 fold and assayed. Activity assay results are shown in Figure 22.

#### Preparation of Corn Seed Fiber

Corn pericarp coarse fiber was isolated by steeping yellow dent #2 kernels for 48hrs at 50 °C in 2000 ppm sodium metabisulfite( (Aldrich). Kernels were mixed with water in equal parts and blended in a Waring laboratory heavy duty blender with the blade in reverse orientation. Blender was controlled with a variable autotransformer (Staco Energy) at 50% voltage output for 2 min. Blended material was washed with tap water over a standard test sieve #7(Fisher scientific) to separate coarse fiber from starch fractions. Coarse fiber and embryos were separated by floating the fiber way from the embryos with hot tap water in a 4L beaker (Fisher scientific). The fiber was then soaked in ethanol prior to drying overnight in a vacuum oven( Precision) at 60° C. Corn coarse fiber derived form corn kernel pericarp was milled with a laboratory mill 3100 fitted with a mill feeder 3170(Perten instruments) to 0.5mm particle size.

#### Corn Fiber Hydrolysis Assay

Course fiber (CF) was suspended in 50 mM citrate-phosphate buffer, pH 5.2 at 30 mg/ 5 ml buffer. The CF stock was vortexed and transferred to a 40 ml modular reservoir (Beckman, Cat. No. 372790). The solution was mixed well then 100 ul transferred to a 96 well plate (Corning Inc., Cat. No.9017, polystyrene, flat bottom). Enzyme was added at 1-10 ul/well and the final volume adjusted to 110 ul with buffer. CF background controls contained 10 ul of buffer only. Plates were sealed with aluminum foil and incubated at 37°C with constant shaking for 18 hours. The plates were centrifuged for 15 min at 4000 rpm. 1-10 ul of CF supernatant was transferred to a 96 well plate preloaded with 100 ul of BCA reagents (BCA-reagents: Reagent A (Pierce, Prod.# 23223), Reagent B (Pierce, Prod.# 23224). The final volume was adjusted to 110 ul. The plate was sealed with aluminum foil and placed at 85°C for 30 min. Following incubation at 85°C, the plate was centrifuged for 5 min at 2500 rpm. Absorbance



values were read at 562 nm (Molecular Devices, Spectramax Plus). Samples were quantified with D-glucose and D-xylose (Sigma) calibration curves. Assay results are reported as total sugar released.

#### Measurement of total sugar released by Ferulic Acid Esterase in Corn Seed Fiber Hydrolysis Assay

Results from the recombinant FAE-1 fiber hydrolysis assay showed no increase in total reducing sugars (data not shown). These results were not unexpected since it has been reported in the literature that an increase in total reducing sugars is detectable only when other hydrolytic enzymes are used in combination with the FAE ( Yu et al J. Agric. Food Chem. 2003, 51, 218-223). Figure 23 shows that addition of FAE-2 to a fungal supernatant which had been grown on corn fiber, shows an increase in total reducing sugars. This suggests that FAE does play an important role in corn fiber hydrolysis.

Figure 23 shows Corn Fiber Hydrolysis assay results showing increase in release of total reducing sugars from corn fiber with addition of FAE-2 to fungal supernatant (FS9).

#### Analysis of Ferulic Acid released from corn seed fiber by FAE-1

FAE activity on corn fiber was tested by following the release of ferulic acid as described in Walfron and Parr (1996) ( Waldron, KW, Parr AJ 1996 Vol 7 pages 305-312 Phytochem Anal) with slight modification. Corn coarse fiber derived from corn kernel pericarp was milled with a laboratory mill 3100 fitted with a mill feeder 3170 (Perten instruments) to 0.5mm particle size and used as substrate at a concentration of 10 mg/ml. 1 ml assays were conducted in 24 well Becton Dickenson Multiwell™. Substrate was incubated in 50 mM citrate phosphate pH 5.4 at 50° C at 110 rpm for 18 hrs in the presence and absence of recombinant FAE. After the incubation period, samples were centrifuged for 10 minutes at 13,000 rpm prior to ethyl acetate extraction. All solvents and acids used were from Fisher Scientific. 0.8 ml of supernatant was acidified with 0.5 ml acetic glacial acid and extracted three times with equivalent volume of

ethyl acetate. Organic fractions were combined and speed vac to dryness (Savant) at 40° C. Samples were then suspended with 100µl of methanol and used for HPLC analysis.

HPLC chromatography was carried out as follows. Ferulic acid (ICN Biomedicals) was used as standard in HPLC analysis (data not shown). HPLC analysis was conducted with a Hewlett Packard series 1100 HPLC system. The procedure employed a C<sub>18</sub> fully capped reverse phase column (XterraRp<sub>18</sub>, 150mm X 3.9mm i.d. 5µm particle size) operated in 1.0 ml min<sup>-1</sup> at 40°C. Ferulic Acid was eluted with a gradient of 25 to 70 % B in 32 min (solvent A: H<sub>2</sub>O, 0.01%b TFA; solvent B: MeCN, 0.0075%).

As shown in Figure 24, FA released from corn fiber was 2-3 fold higher than control when treated with 10 or 100 ul of FAE-1. These results clearly show that FAE-1 is capable of hydrolyzing corn fiber.

#### Example 54

##### Functionality in fermentation of maize expressed glucoamylase and amylase

This example demonstrates that maize-expressed enzymes will support fermentation of starch in a corn slurry in the absence of added enzyme and without cooking the corn slurry. Maize kernels that contain *Rhizopus oryzae* glucoamylase (ROGA) (SEQ ID NO: 49) were produced as described in Example 32. Maize kernels that contain the barley low-pI α-amylase (AMYI) (SEQ ID NO: 88) are produced as described in Example 46. The following materials are used in this example:

*Aspergillus niger* glucoamylase (ANGA) was purchased from Sigma.

*Rhizopus* species glucoamylase (RxGA) was purchased from Wako as a dry crystalline powder and made up in 10 mM NaAcetate pH 5.2, 5 mM CaCl<sub>2</sub>. at 10 mg/ml.

MAMYI Microbially produced AMYI was prepared at approximately 0.25 mg/ml in 10 mM NaAcetate pH 5.2, 5 mM CaCl<sub>2</sub>.

Yeast was *Saccharomyces cereviceae*

YE was a sterile 5% solution of yeast extract in water

Yeast starter contained 50 g maltodextrin, 1.5 g yeast extract, 0.2 mg ZnSO<sub>4</sub> in a total volume of 300 ml of water. the medium was sterilized by autoclaving after preparation. After cooling to room temperature, 1 ml of tetracycline (10 mg/ml in ethanol), 100 µl AMG300 glucoamylase and 155 mg active dry yeast. were added. The mixture was then shaken at 30 °C for 22 h. The overnight yeast culture was diluted 1/10 with water and A600 measured to determine the yeast number, as described in Current Protocols in Molecular Biology.

ROGA flour Kernels were pooled from several T0 lines shown to have active glucoamylase. The seeds were ground in the Kleco, and all flour was pooled.

AMYI flour Kernels from T0 corn expressing AMYI were pooled and ground as above.

Control flour Kernels from with similar genetic background were ground in the same fashion as the ROGA expressing corn

An inoculation mixture was prepared in a sterile tube; it contained per 1.65 ml: yeast cells ( $1 \times 10^7$ ), yeast extract (8.6 mg), tetracycline (55 µg). 1.65 ml was added / g flour to each fermentation tube.

Fermentation preparation: Flour was weighed out at 1.8 g / tube into tared 17 x 100 mm sterile polypropylene. 50 µl of 0.9 M H<sub>2</sub>SO<sub>4</sub> was added to bring the final pH prior to fermentation to 5. The inoculation mixture (2.1 ml) was added / tube. along with RXGA, AMYI-P and amylase desalting buffer as indicated below. The quantity of buffer was adjusted based on moisture content of each flour so that the total solids content was constant in each tube. The tubes were mixed thoroughly, weighed and placed into a plastic bag and incubated at 30 °C.

Table 21

Tube	Flours			Innoculation	Microbial enzymes		Amylase desalting Buffer
	Control	ROGA	AMYI		RXGA	AMYI-P	
	g	g	g	ml	ml	ml	ml
A	1.8			2.1	0	0	

B	1.8			2.1	0.036	0	1
C	1.8			2.1	0.036	1	0
D	1.8			2.1	0	1	0.036
E	1.6		0.2	2.1	0.036	0	1
F	0.2	1.6		2.1			1
G	0.2	1.6		2.1	0	1	0
H	0	1.6	0.2	2.1		0	1

The fermentation tubes were weighed at intervals over the 67 h time course. Loss of weight corresponds to evolution of CO<sub>2</sub> during fermentation. The ethanol content of the samples was determined after 67 h of fermentation by the DCL ethanol assay method. The kit (catalogue # 229-29) was purchased from Diagnostic Chemicals Limited, Charlottetown, PE, Canada, D1E 1B0. Samples (10 µl) were drawn in triplicate from each fermentation tube and diluted into 990 µl of water. 10 µl of the diluted samples were mixed with 1.25 ml of a 12.5/1 mixture of assay buffer / ADH-NAD reagent. Standards (0, 5, 10, 15 & 20% v/v ETOH) were diluted and assayed in parallel. Reactions were incubated at 37 °C for 10 min, then A340 read. Standards were prepared in duplicate, samples from each fermentation were prepared in triplicate (including the initial dilution). The weight of the samples changed with time as detailed in table below. The weight loss is expressed as a percentage of the initial sample weight at time 0.

Table 22

		Time (h)					
		0	18	24	42	48	67
Sample	Flour Composition	% wgt loss					
A	Control	0.00	8.09	9.38	12.96	13.83	16.85
B	Control + RXGA	0.00	11.48	14.20	21.79	23.83	24.63
C	Control + RXGA + MAMYI	0.00	17.90	23.27	36.48	39.07	47.59
D	Control + MAMYI	0.00	13.70	17.72	28.27	30.80	38.27
E	Control +RXGA + AMYI flour	0.00	16.85	21.60	33.95	36.98	45.74
F	ROGA flour	0.00	9.81	11.74	16.96	18.39	23.17
G	ROGA flour + MAMYI	0.00	15.53	19.69	29.75	32.11	39.94
H	ROGA flour + AMYI flour	0.00	13.35	16.27	23.60	25.53	31.68

These data show that the ROGA enzyme expressed in maize increases fermentation rate as compared to the no-enzyme control. It also confirms previous data indicating that the AMYI

enzyme expressed in maize kernels is a potent activator of fermentation of the starch in corn. The ethanol contents are detailed below.

Table 23

Sample	Flour Composition	ETOH % v/v	Standard deviation
A	Control	2.09	0.08
B	Control + RXGA	7.97	0.18
C	Control + RXGA + MAMYI	13.47	0.27
D	Control + MAMYI	11.26	0.12
E	Control +RXGA + AMYI flour	12.28	0.08
F	ROGA flour	3.55	0.05
G	ROGA flour + MAMYI	11.29	0.18
H	ROGA flour + AMYI flour	8.58	0.13

These data also demonstrate that expressing *Rhizopus oryzae* glucoamylase in maize facilitates increased fermentation of the starch in corn. Similarly, expression of the barley amylase in maize makes corn starch more fermentable with out adding exogenous enzymes.

#### **Example 55** **Cellobiohydrolase I**

The *Trichoderma reesei* cellobiohydrolase I (CBH I) gene was amplified and cloned by RT-PCR based on a published database sequence (accession # E00389). The cDNA sequence was analyzed for the presence of a signal sequence using the SignalP program, which predicted a 17 amino acid signal sequence. The DNA sequence encoding the signal sequence was replaced with an ATG by PCR, as shown in the sequence (SEQ ID NO: 79). This cDNA sequence was used to make subsequent constructs. Additional constructs are made by substituting a maize optimised version of the gene (SEQ ID NO: 93).

#### **Example 56** **Cellobiohydrolase II**

The *Trichoderma reesei* cellobiohydrolase II (CBH II) gene was amplified and cloned by RT-PCR based on a published database sequence (accession # M55080). The cDNA sequence was analyzed for the presence of a signal sequence using the SignalP program, which predicted

an 18 amino acid signal sequence. The DNA sequence encoding the signal sequence was replaced with an ATG by PCR, as shown in the sequence (SEQ ID NO: 81). This cDNA sequence was used to make subsequent constructs. Additional constructs are made by substituting a maize optimised version (SEQ ID NO: 94) of the gene.

### Example 57

#### Construction of transformation vectors for the *Trichoderma reesii* cellobiohydrolase I and cellobiohydrolase II

Cloning of the *Trichoderma reesii* cellobiohydrolase I (*cbhi*) cDNA without the native N-terminal signal sequence is described in Example 55. Expression cassettes were constructed to express the *Trichoderma reesii* cellobiohydrolase I cDNA in maize endosperm with various targeting signals as follows:

Plasmid 12392 comprises the *Trichoderma reesii cbhi* cDNA cloned behind the  $\gamma$  zein promoter for expression specifically in the endosperm for expression in the cytoplasm.

Plasmid 12391 comprises the maize  $\gamma$ -zein N-terminal signal sequence (MRVLLVALALLALAASATS)(SEQ ID NO:17) fused to *Trichoderma reesii cbhi* cDNA as described above in Example 1 for targeting to the endoplasmic reticulum and secretion into the apoplast (Torrent et al. 1997). The fusion was cloned behind the  $\gamma$  zein promoter for expression specifically in the endosperm.

Plasmid 12392 comprises the  $\gamma$ -zein N-terminal signal sequence fused to the *Trichoderma reesii cbhi* cDNA with a C-terminal addition of the sequence KDEL for targeting to and retention in the endoplasmic reticulum (ER) (Munro and Pelham, 1987). The fusion was cloned behind the maize  $\gamma$  zein promoter for expression specifically in the endosperm.

Plasmid 12656 comprises the waxy amyloplast targeting peptide (Klosgen et al., 1986) fused to the *Trichoderma reesii cbhi* cDNA for targeting to the amyloplast. The fusion was cloned behind the maize  $\gamma$  zein promoter for expression specifically in the endosperm.

All expression cassettes were moved into a binary vector (pNOV2117) for transformation into maize via *Agrobacterium* infection. The binary vector contained the phosphomannose

isomerase (PMI) gene which allows for selection of transgenic cells with mannose. Transformed maize plants were either self-pollinated or outcrossed and seed was collected for analysis.

Additional constructs (plasmids 12652, 12653, 12654 and 12655) were made with the targeting signals described above fused to *Trichoderma reesii* cellobiohydrolaseII (*cbhii*) cDNA in precisely the same manner as described for the *Trichoderma reesii* *cbhi* cDNA. These fusions were cloned behind the maize Q protein promoter (50Kd  $\gamma$  zein) (SEQ ID NO: 98) for expression specifically in the endosperm and transformed into maize as described above. Transformed maize plants were either self-pollinated or outcrossed and seed was collected for analysis.

Combinations of the enzymes can be produced either by crossing plants expressing the individual enzymes or by cloning several expression cassettes into the same binary vector to enable co-transformation.

### **Example 58**

#### **Expression of a Cbhi in corn**

T1 seed from self-pollinated maize plants transformed with either plasmid 12390, 12391 or 12392 was obtained. The 12390 construct targets the expression of the CbhI in the endoplasmic reticulum of the endosperm, the 12391 construct targets the expression of the CbhI in the apoplast of the endosperm and the 12392 construct targets the expression of the CbhI in the cytoplasm of the endosperm.

**Extraction and detection of the CbhI from corn-flour:** Polyclonal antibodies to CbhI and CbhII were produced in goat according to established protocols. Flour from the CbhI transgenic seeds was obtained by grinding them in an Autogizer grinder. Approximately 50 mg of flour was resuspended in 0.5ml of 20mM NaPO<sub>4</sub> buffer (pH 7.4), 150mM NaCl followed by incubation for 15 minutes at RT with continuous shaking. The incubated mixture was then spun for 10min. at 10,000xg. The supernatant was used as enzyme source. 30  $\mu$ l of this extract was loaded on a 4-12 % NuPAGE gel (invitrogen) and separated in the NuPAGE MES running buffer (invitrogen). Protein was blotted onto nitrocellulose membranes and Western blot

analysis was done following established protocols using the specific antibodies described above followed by alkaline phosphatase conjugated rabbit antigoat IgG (H+L) . Alkaline phosphatase activity was detected by incubation of the membranes with ready to use BCIP/MBT (plus) substrate from Moss Inc.

Western Blot analysis was done of T1 seeds from different events transformed with plasmid 12390. Expression of Cbhl protein was compared to the non-transgenic control, and was detected in a number of events.

The Cracked Corn Assay was performed essentially as described in Example 49, using transgenic seed expressing Cbhl. Starch recovery from the transgenic seed was measured and the results are set forth in Table 24.

Table 24.

Conditions	Line 3-non expressing control	Line 4- CBHI expressing
	Starch (mg)	
400ppm SO <sub>2</sub> -No Bromelain	40.2	78.1
400ppmSO <sub>2</sub> -Plus Bromelain	48.1	118.7
2000ppm SO <sub>2</sub> -No Bromelain	47.5	73.1
2000ppmSO <sub>2</sub> -Plus Bromelain	49.2	109

### Example 59

#### Preparation of Endoglucanase I Constructs

A *Trichoderma reesei* endoglucanase I (EGLI) gene was amplified and cloned by PCR based on a published database sequence (Accession # M15665; Penttila et al., 1986). Because only genomic sequences could be obtained, the cDNA was generated from the genomic sequence by removing 2 introns using Overlap PCR. The resulting cDNA sequence was analyzed for the presence of a signal sequence using the SignalP program, which predicted a 22 amino acid signal sequence. The DNA sequence encoding the signal sequence was replaced with an ATG by PCR, as shown in the sequence (SEQ ID NO: 83). This cDNA sequence was used to make subsequent constructs as set forth below.



### Overlap PCR

Overlap PCR is a technique (Ho et al., 1989) used to fuse complementary ends of two or more PCR products, and can be used to make base pair (bp) changes, add bp, or delete bp. At the site of the intended bp change, forward and reverse mutagenic primers (Mut-F and Mut-R) are made that contain the intended change and 15 bp of sequence on either side of the change. For example, to remove an intron, the primers would consist of the final 15 bp of exon 1 fused to the first 15 bp of exon 2. Primers are also prepared that anneal to the ends of the sequence to be amplified, e.g. ATG and STOP codon primers. PCR amplification of the products proceeds with the ATG/Mut-R primer pair and the Mut-F/STOP primer pair in independent reactions. The products are gel purified and fused together in a PCR without added primers. The fusion reaction is separated on a gel, and the band of the correct size is gel purified and cloned. Multiple changes can be accomplished simultaneously through the addition of additional mutagenic primer pairs.

### EGLI Plant Expression Constructs

Expression cassettes were made to express the *Trichoderma reesei* EGLI cDNA in maize endosperm as follows:

**13025** comprises the *T. reesei* EGLI gene cloned behind the maize  $\gamma$ -zein promoter for cytoplasmic localization and expression specifically in the endosperm.

**13026** comprises the maize  $\gamma$ -zein N-terminal signal peptide (MRVLLVALALLALAASATS) fused to the *T. reesei* EGLI gene for targeting to the endoplasmic reticulum and secretion into the apoplast. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

**13027** comprises the maize  $\gamma$ -zein N-terminal signal peptide fused to the *T. reesei* EGLI gene with a C-terminal addition of the sequence KDEL for targeting to and retention in the endoplasmic reticulum. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

**13028** comprises the maize Granule Bound Starch Synthase I (GBSSI) N-terminal signal peptide (N-terminal 77 amino acids) fused to the *T. reesei* EGLI gene for targeting to the lumen of the amyloplast. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

**13029** comprises the maize GBSSI N-terminal signal peptide fused to the *T. reesei* EGLI gene with a C-terminal addition of the starch binding domain (C-terminal 301 amino acids) of the maize GBSSI gene for targeting to the starch granule. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

Additional Expression cassettes are generated using a maize optimised version of EGLI (SEQ ID NO: 95)

#### **EGLI Enzyme Assays**

EGLI enzyme activity is measured in maize transgenics using the Malt Beta-Glucanase Assay Kit (Cat # K-MBGL) (Megazyme International Ireland Ltd.) The enzymatic activity of EGL I expressors is tested in the Corn Fiber Hydrolysis Assay as described in Example 53.

**Example 60**  
 **$\beta$ -Glucosidase 2**

A *Trichoderma reesei*  $\beta$ -Glucosidase 2 (BGL2) gene was amplified and cloned by RT-PCR based on sequence Accession # AB003110 (Takashima et al., 1999).

**BGL2 Plant Expression Constructs**

Expression cassettes were made to express the *Trichoderma reesei* BGL2 cDNA (SEQ ID NO: 89) in maize endosperm as follows:

**13030** comprises the *T. reesei* BGL2 gene cloned behind the maize  $\gamma$ -zein promoter for cytoplasmic localization and expression specifically in the endosperm.

**13031** comprises the maize  $\gamma$ -zein N-terminal signal peptide (MRVLLVALALLALAASATS) fused to the *T. reesei* BGL2 gene for targeting to the endoplasmic reticulum and secretion into the apoplast. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

**13032** comprises the maize  $\gamma$ -zein N-terminal signal peptide fused to the *T. reesei* BGL2 gene with a C-terminal addition of the sequence KDEL for targeting to and retention in the endoplasmic reticulum. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

**13033** comprises the maize Granule Bound Starch Synthase I (GBSSI) N-terminal signal peptide (N-terminal 77 amino acids) fused to the *T. reesei* BGL2 gene for targeting to the lumen of the amyloplast. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

**13034** comprises the maize GBSSI N-terminal signal peptide fused to the *T. reesei* BGL2 gene with a C-terminal addition of the starch binding domain (C-terminal 301 amino acids) of the

maize GBSSI gene for targeting to the starch granule. The fusion was cloned behind the maize  $\gamma$ -zein promoter for expression specifically in the endosperm.

Additional Expression cassettes are generated by substituting a maize optimized version of BGL2 (SEQ ID NO: 96).

All expression cassettes are inserted into the binary vector pNOV2117 for transformation into maize via *Agrobacterium* infection. The binary vector contained the phosphomannose isomerase (PMI) gene which allows for selection of transgenic cells with mannose. Transformed maize plants were either self-pollinated or outcrossed and seed was collected for analysis.

#### BGL2 Enzyme Assays

BGL2 enzyme activity is measured in transgenic maize using a protocol modified from Bauer and Kelly (Bauer, M.W. and Kelly, R.M. 1998. The family 1  $\beta$ -glucosidases from *Pyrococcus furiosus* and *Agrobacterium faecalis* share a common catalytic mechanism. Biochemistry 37: 17170-17178). The protocol can be modified to incubate samples at 37°C instead of 100°C. The enzymatic activity of BGL2-expressors is tested in the Fiber Hydrolysis Assay.

**Example 61** **$\beta$ -Glucosidase D**

The *Trichoderma reesei*  $\beta$ -Glucosidase D (CEL3D) gene was amplified and cloned by PCR based on a published database sequence (accession # AY281378; Foreman et al., 2003). Because only genomic sequences could be obtained, the cDNA was generated from the genomic sequence by removing an intron using Overlap PCR, as described in Example 58. The resulting cDNA (SEQ ID NO: 91) may be used for subsequent constructs. A maize optimised version (SEQ ID NO: 97) of the resulting cDNA may also be used for constructs.

Plant constructs can be generated and  $\beta$ -glucosidase assays can be performed as described for BGL2 in Example 60, replacing BGL2 with CEL3D.

**Example 62****Lipases**

cDNAs encoding lipases are generated using sequences from Accession # D85895, AF04488, and AF04489 (Tsuchiya et al., 1996; Yu et al., 2003) and methodology set forth in Examples 59-60.

Lipase enzyme activity can be measured in transgenic maize using the Fluorescent Lipase Assay Kit (Cat # M0612)(Marker Gene Technologies, Inc.). Lipase activity can also be measured *in vivo* using the fluorescent substrate 1,2-dioleoyl-3-(pyren-1-yl)decanoyl-*rac* glycerol (M0258), also from Marker Gene Technologies, Inc.

**Example 63****Expression of Phytase in Rice**

Vectors 11267 and 11268 comprise binary vectors that encode Nov9x phytase. Expression of the Nov9x phytase gene in both vectors is under the control of the rice glutelin-1 promoter (SEQ ID NO:67). Vectors 11267 and 11268 are derived from pNOV2117.

The Nov9x phytase expression cassette in vector 11267 comprises the rice glutelin-1 promoter, the Nov9x phytase gene with apoplast targeting signal, a PEPC intron, and the 35S terminator. The product of the Nov9x phytase coding sequence in vector 11267 is shown in SEQ ID NO: 110 .

The Nov9x phytase expression cassette in vector 11268 comprises the rice glutelin-1 promoter, the Nov9x phytase gene with ER retention (SEQ ID NO:111), a PEPC intron, and the 35S terminator. The product of the Nov9x phytase coding sequence in vector 11268 is shown in SEQ ID NO: 112.

**11267 Nov9x phytase with apoplast targeting DNA sequence (SEQ ID NO: 109).  
Translation start and stop codons are underlined. The sequence encoding the signal  
sequence of the 27-kD gamma-zein protein is in bold.**

atgagggtgtgtcgttgcctcgtctcgtctcgtcgcgagcgccaccagcgctgcgcagtcgagccggagctgaagctgg  
agtcctgtgtgatcgtgtccgccacggcgctgcgcgccccgaccaaggccaccagctcatgcaggacgtgacccggagcctggcc  
gacctggccggtgaagctcggcgagctgacccgcgcggcgagctgatcgctacctcggccactactggccagcgcctcgtg  
gccgacggcctcctccgaagtgcggctgccgcagtcggccaggtggccatcatcgccgacgtggacgagcgcacccgcaagacc  
ggcgaggccttcgccggcgccctcgcggcgactgcgccatcaccgtgcacacccaggccgacacctcctccccggaccgctcttcaa  
cccgtcaagaccggcggtgtgccagctcgacaacgccaacgtgaccgacgccatcctggagcgcgcggcggtctcatcgccgacttc  
accggccactaccagaccgccttcgcgagctggagcgcgtgtcctaacttcccgagtcacaacctctgcttcaagcgcgagaagcagga  
cgagtcctgctcctcaccagccctcccgctccgagctgaaggtgtccgccgactgcgtgtcctcaccggcgccgtgtcctcgcctcc  
atgctaccggaatcttctcctccagcaggccacgggcatgccggagccgggtggggccgcatcaccgactccaccagtggaacac  
cctcctcctcccaacgcccagttcgacctcctccagcgcacccggaggtggccgctcccgccaccccgctcctcgacctatc  
aagaccgccctaccccgccacccggcgagagcagggcctacggcgtagacctcccgacctccgtgtcttcatcgccggccacgacac  
caacctcgccaaacctcggcgccgcttgagctgaactggacctcccgggccagccgacaacaccccgccggcgccgagctggt  
gttcgagcgctggcgccgcctcctccgacaactcccagtggaattcaggtgtccctcgtgtccagacctccagcagatgcgcgacaagacc  
ccgctcctcctcaacaccccgccggcgaggtgaagctcaccctcgcggctgcgagggagcgaacgccagggcatgtgtcctcgtg  
ccggcttaccagatcgtgaacgagggccgcacccggcctgtcctcctaa

**11267 Nov9x phytase with apoplast targeting gene product (SEQ ID NO:110). The signal  
sequence of the 27-kD gamma-zein protein is in bold.**

mrllvalallalaasatsaaqsepelklesvvihsrhgvraptkqlmqdvtpdawptwvklgltpggeliaylghywrqlva  
dglpkcgcpqsgqvaiiadvertrktgeafaaglapdcaivhtqadtssdpdlnplktgvcqldnanvtdaileraggsiadfghy  
qtafrelevlnfpqsnclckrekqdescsltqalpselkvsadcvsltgavslasmlteifllqqagmppegwgritdshqwnllslhn  
aqfdllqrpevarsratplldliktaltphppqkaygvltptsvlfiaghdtnlanlgalelnwtlpgqpdntppggelvferwrslsdn  
sqwivqslvfqlqqmrdrktplslntppgevkllagceernaagmcslagftqivnearipacsl

**11268 Nov9x phytase with ER retention DNA sequence (SEQ ID NO:111). The sequence  
encoding the signal sequence of the 27-kD gamma-zein protein is in bold. The sequence  
encoding the SEKDEL hexapeptide ER retention signal is underlined.**

atgagggtgtgtcgttgcctcgtctcgtctcgtcgcgagcgccaccagcgctgcgcagtcgagccggagctgaagctgg  
agtcctgtgtgatcgtgtccgccacggcgctgcgcgccccgaccaaggccaccagctcatgcaggacgtgacccggagcctggcc  
gacctggccggtgaagctcggcgagctgacccgcgcggcgagctgatcgctacctcggccactactggccagcgcctcgtg  
gccgacggcctcctccgaagtgcggctgccgcagtcggccaggtggccatcatcgccgacgtggacgagcgcacccgcaagacc  
ggcgaggccttcgccggcgccctcgcggcgactgcgccatcaccgtgcacacccaggccgacacctcctccccggaccgctcttcaa  
cccgtcaagaccggcggtgtgccagctcgacaacgccaacgtgaccgacgccatcctggagcgcgcggcggtctcatcgccgacttc  
accggccactaccagaccgccttcgcgagctggagcgcgtgtcctaacttcccgagtcacaacctctgcttcaagcgcgagaagcagga  
cgagtcctgctcctcaccagccctcccgctccgagctgaaggtgtccgccgactgcgtgtcctcaccggcgccgtgtcctcgcctcc  
atgctaccggaatcttctcctccagcaggccacgggcatgccggagccgggtggggccgcatcaccgactccaccagtggaacac  
cctcctcctcccaacgcccagttcgacctcctccagcgcacccggaggtggccgctcccgccaccccgctcctcgacctatc  
aagaccgccctaccccgccacccggcgagagcagggcctacggcgtagacctcccgacctccgtgtcttcatcgccggccacgacac  
caacctcgccaaacctcggcgccgcttgagctgaactggacctcccgggccagccggacaacaccccgccggcgccgagctggt  
gttcgagcgctggcgccgcctcctccgacaactcccagtggaattcaggtgtccctcgtgtccagacctccagcagatgcgcgacaagacc

ccgctctccctcaacacccccgccggcgaggtgaagctcaccctgccggctgcgaggagcgcaacgccagggcatgtctccctcg  
ccggcttcaccagatcgtgaacgaggcccgcatcccgctcctcctcccgagaaggacgagctgtaa

**11268 Nov9x phytase with ER retention, gene product (SEQ ID NO: 112). The signal sequence of the 27-kD gamma-zein protein is in bold. The ER retention signal is underlined.**

**mrvtllvalallalaasatsaaqsepelklesvvivsrhgvraptkatqlmqdvtpdawptwvpklgeltprggeliaylghywrqlva**  
**dgllpkcgcpqsgqvaiiadvdertrktgeafaaglapdcaivhtqadtsspdplfnplktgvcqldnanvtdaileraggsiadftghy**  
**qtafrelevlnfpqsnlclkrekdscsltqalpselkvsadcvsltgavslasmlteifllqqaqgmpepgwgritdshqwnllslhn**  
**aqfdllrtpevarsratplldiktaltphppqkqaygvltptsvlfiaghdmlanlggalelnwtlpgqpdpntppggelvferrwrlsdn**  
**sqwiqvslvfqtlqqmrktpslntppgevkltlagceernaqgmcslagftqivnearipacslsekdel**

#### Generation of transgenic rice plants

Rice (*Oryza sativa*) is used for generating transgenic plants. Various rice cultivars can be used (Hiei et al., 1994, *Plant Journal* 6:271-282; Dong et al., 1996, *Molecular Breeding* 2:267-276; Hiei et al., 1997, *Plant Molecular Biology*, 35:205-218). Also, the various media constituents described below may be either varied in concentration or substituted. Embryogenic responses are initiated and/or cultures are established from mature embryos by culturing on MS-CIM medium (MS basal salts, 4.3 g/liter; B5 vitamins (200 x), 5 ml/liter; Sucrose, 30 g/liter; proline, 500 mg/liter; glutamine, 500 mg/liter; casein hydrolysate, 300 mg/liter; 2,4-D (1 mg/ml), 2 ml/liter; adjust pH to 5.8 with 1 N KOH; Phytigel, 3 g/liter). Either mature embryos at the initial stages of culture response or established culture lines are inoculated and co-cultivated with the *Agrobacterium* strain LBA4404 containing the desired vector construction. *Agrobacterium* is cultured from glycerol stocks on solid YPC medium (100 mg/L spectinomycin and any other appropriate antibiotic) for ~2 days at 28 °C. *Agrobacterium* is re-suspended in liquid MS-CIM medium. The *Agrobacterium* culture is diluted to an OD600 of 0.2-0.3 and acetosyringone is added to a final concentration of 200 µM. *Agrobacterium* is induced with acetosyringone before mixing the solution with the rice cultures. For inoculation, the cultures are immersed in the bacterial suspension. The liquid bacterial suspension is removed and the inoculated cultures are placed on co-cultivation medium and incubated at 22°C for two days. The cultures are then transferred to MS-CIM medium with Ticarcillin (400 mg/liter) to inhibit the growth of *Agrobacterium*. For constructs utilizing the PMI selectable marker gene (Reed et al., *In Vitro Cell. Dev. Biol.-Plant* 37:127-132), cultures are transferred to selection medium containing



Mannose as a carbohydrate source (MS with 2% Mannose, 300 mg/liter Ticarcillin) after 7 days, and cultured for 3-4 weeks in the dark. Resistant colonies are then transferred to regeneration induction medium (MS with no 2,4-D, 0.5 mg/liter IAA, 1 mg/liter zeatin, 200 mg/liter Ticarcillin 2% Mannose and 3% Sorbitol) and grown in the dark for 14 days. Proliferating colonies are then transferred to another round of regeneration induction media and moved to the light growth room. Regenerated shoots are transferred to GA7-1 medium (MS with no hormones and 2% Sorbitol) for 2 weeks and then moved to the greenhouse when they are large enough and have adequate roots. Plants are transplanted to soil in the greenhouse and grown to maturity.

#### Example 64

##### Analysis of Transgenic Rice Seed Expressing Nov9X Phytase

##### ELISA For The Quantitation Of Nov9X Phytase From Rice Seed

Quantitation of phytase expressed in transgenic rice seed was assayed by ELISA. One (1g) rice seed was ground to flour in a Kleco seed grinder. 50 mg of flour was resuspended in the sodium acetate buffer described in example – for Nov9X phytase activity assay and diluted as required for the immunoassay. The Nov9X immunoassay is a quantitative sandwich assay for the detection of phytase that employs two polyclonal antibodies. The rabbit antibody was purified using protein A, and the goat antibody was immunoaffinity purified against recombinant phytase (Nov9X) protein produced in *E.coli* inclusion bodies. Using these highly specific antibodies, the assay can measure picogram levels of phytase in transgenic plants. There are three basic parts to the assay. The phytase protein in the sample is captured onto the solid phase microtiter well using the rabbit antibody. Then a “sandwich” is formed between the solid phase antibody, the phytase protein, and the secondary antibody that has been added to the well. After a wash step, where unbound secondary antibody has been removed, the bound antibody is detected using an alkaline phosphatase-labeled antibody. Substrate for the enzyme is added and color development is measured by reading the absorbance of each well. The standard curve uses a four-parameter curve fit to plot the concentrations versus the absorbance.

##### Phytase activity assay

Determination of phytase activity, based upon the estimation of inorganic phosphate released on hydrolysis of phytic acid, can be performed at 37°C following the method of Engelen, A.J. et al., *J. AOAC, Inter.*, 84, 629 (2001). One unit of enzyme activity is defined as the amount of enzyme that liberates 1  $\mu\text{mol}$  of inorganic phosphate per minute under assay conditions. For example, phytase activity may be measured by incubating 2.0 ml of the enzyme preparation with 4.0 ml of 9.1 mM sodium phytate in 250 mM sodium acetate buffer pH 5.5, supplemented with 1 mM  $\text{CaCl}_2$  for 60 minutes at 37°C. After incubation, the reaction is stopped by adding 4.0 ml of a color-stop reagent consisting of equal parts of a 10% (w/v) ammonium molybdate and a 0.235% (w/v) ammonium vanadate stock solution. Precipitate is removed by centrifugation, and phosphate released is measured against a set of phosphate standards spectrophotometrically at 415 nm. Phytase activity is calculated by interpolating the A415 nm absorbance values obtained for phytase containing samples using the generated phosphate standard curve.

This procedure may be scaled down to accommodate smaller volumes and adapted to preferred containers. Preferred containers include glass test tubes and plastic microplates. Partial submersion of the reaction vessel(s) in a water bath is essential to maintain constant temperature during the enzyme reaction.

Table 24

Trans-genic line	$\mu\text{g}$ phytase/g flour*	Phytase activity units per g flour**	Endogenous inorganic phosphate released by cooking of dehusked rice seed ( $\mu\text{mol/gseed}$ )	Endogenous inorganic phosphate released by cooking of dehusked, polished rice seed ( $\mu\text{mol/gseed}$ )
Wild type	0	0	1.442	0.469
1	510	916	1.934	0.840
2	1518	2800	2.894	1.073

\* $\mu\text{g}$  phytase was assayed by a sandwich ELISA

\*\*Phytase activity was assayed by Phytase activity assay as described above.

#### Assay of Inorganic Phosphate Release During Cooking of Transgenic Rice Expressing Phytase

Two samples of 1g seed from selected rice transgenic lines and a control wildtype line was dehusked using a benchtop Kett TR200 automatic rice husker. One sample was then

polished for 30 seconds in a Kett Rice polisher. Two volumes of H<sub>2</sub>O was added to each sample and the rice was cooked by immersing the tubes into a beaker of water. The water was brought to a boil and held in a full rolling boil for 10 minutes. The "cooked" rice seed was then ground to a paste with water bringing the total volume of the slurry to 6 ml. The slurry was centrifuged at 15,000xg for 10 minutes and the clear supernatant assayed for released endogenous inorganic phosphate. The assay of released phosphate is based on color formation as a result of molybdate and vanadate ions complexing with inorganic phosphate and is measured spectrophotometrically at 415nm as described in example – for phytase enzymatic activity. The results are in Table 24.

All publications, patents and patent applications are incorporated herein by reference. While in the foregoing specification this invention has been described in relation to certain preferred embodiments thereof, and many details have been set forth for purposes of illustration, it will be apparent to those skilled in the art that the invention is susceptible to additional embodiments and that certain of the details described herein may be varied considerably without departing from the basic principles of the invention.

#### SEQUENCE LISTING

<110> Lanahan, Mike  
<120> Self-processing Plants and Plant Parts  
<130> 109846.317  
<140> US 60/315,281  
<141> 2001-08-27  
<160> 60  
<170> FastSEQ for Windows Version 4.0  
<210> 1  
<211> 436  
<212> PRT  
<213> Artificial Sequence  
<220>  
<223> synthetic  
<400> 1

Met Ala Lys Tyr Leu Glu Leu Glu Glu Gly Gly Val Ile Met Gln Ala  
 1 5 10 15  
 Phe Tyr Trp Asp Val Pro Ser Gly Gly Ile Trp Trp Asp Thr Ile Arg  
 20 25 30  
 Gln Lys Ile Pro Glu Trp Tyr Asp Ala Gly Ile Ser Ala Ile Trp Ile  
 35 40 45  
 Pro Pro Ala Ser Lys Gly Met Ser Gly Gly Tyr Ser Met Gly Tyr Asp  
 50 55 60  
 Pro Tyr Asp Tyr Phe Asp Leu Gly Glu Tyr Tyr Gln Lys Gly Thr Val  
 65 70 75 80  
 Glu Thr Arg Phe Gly Ser Lys Gln Glu Leu Ile Asn Met Ile Asn Thr  
 85 90 95  
 Ala His Ala Tyr Gly Ile Lys Val Ile Ala Asp Ile Val Ile Asn His  
 100 105 110  
 Arg Ala Gly Gly Asp Leu Glu Trp Asn Pro Phe Val Gly Asp Tyr Thr  
 115 120 125  
 Trp Thr Asp Phe Ser Lys Val Ala Ser Gly Lys Tyr Thr Ala Asn Tyr  
 130 135 140  
 Leu Asp Phe His Pro Asn Glu Leu His Ala Gly Asp Ser Gly Thr Phe  
 145 150 155 160  
 Gly Gly Tyr Pro Asp Ile Cys His Asp Lys Ser Trp Asp Gln Tyr Trp  
 165 170 175  
 Leu Trp Ala Ser Gln Glu Ser Tyr Ala Ala Tyr Leu Arg Ser Ile Gly  
 180 185 190  
 Ile Asp Ala Trp Arg Phe Asp Tyr Val Lys Gly Tyr Gly Ala Trp Val  
 195 200 205  
 Val Lys Asp Trp Leu Asn Trp Trp Gly Gly Trp Ala Val Gly Glu Tyr  
 210 215 220  
 Trp Asp Thr Asn Val Asp Ala Leu Leu Asn Trp Ala Tyr Ser Ser Gly  
 225 230 235 240  
 Ala Lys Val Phe Asp Phe Pro Leu Tyr Tyr Lys Met Asp Ala Ala Phe  
 245 250 255  
 Asp Asn Lys Asn Ile Pro Ala Leu Val Glu Ala Leu Lys Asn Gly Gly  
 260 265 270  
 Thr Val Val Ser Arg Asp Pro Phe Lys Ala Val Thr Phe Val Ala Asn  
 275 280 285  
 His Asp Thr Asp Ile Ile Trp Asn Lys Tyr Pro Ala Tyr Ala Phe Ile  
 290 295 300  
 Leu Thr Tyr Glu Gly Gln Pro Thr Ile Phe Tyr Arg Asp Tyr Glu Glu  
 305 310 315 320  
 Trp Leu Asn Lys Asp Lys Leu Lys Asn Leu Ile Trp Ile His Asp Asn  
 325 330 335  
 Leu Ala Gly Gly Ser Thr Ser Ile Val Tyr Tyr Asp Ser Asp Glu Met  
 340 345 350  
 Ile Phe Val Arg Asn Gly Tyr Gly Ser Lys Pro Gly Leu Ile Thr Tyr  
 355 360 365  
 Ile Asn Leu Gly Ser Ser Lys Val Gly Arg Trp Val Tyr Val Pro Lys  
 370 375 380  
 Phe Ala Gly Ala Cys Ile His Glu Tyr Thr Gly Asn Leu Gly Gly Trp  
 385 390 395 400  
 Val Asp Lys Tyr Val Tyr Ser Ser Gly Trp Val Tyr Leu Glu Ala Pro  
 405 410 415  
 Ala Tyr Asp Pro Ala Asn Gly Gln Tyr Gly Tyr Ser Val Trp Ser Tyr  
 420 425 430

Cys Gly Val Gly  
435

<210> 2  
<211> 1308  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 2  
atggccaagt acctggagct ggaggagggc ggcgtgatca tgcaggcgtt ctactgggac 60  
gtcccagagcg gaggcattctg gtgggacacc atccgccaga agatccccga gtggtacgac 120  
gccggcatct ccgcgatctg gataccgccca gcttccaagg gcatgtccgg gggctactcg 180  
atgggctacg acccgtagca ctacttcgac ctccggcagat actaccagaa gggcacggcg 240  
gagacgcgct tcgggtccaa gcaggagctc atcaacatga tcaacacggc gcacgcctac 300  
ggcatcaagg tcatcgcgga catcgtgatc aaccacaggg ccggcggcga cctggagtgg 360  
aacccgttcg tcggcgacta cacctggacg gacttctcca aggtcgcctc cggcaagtac 420  
accgccaact acctcgactt ccaccccaac gagctgcacg cgggcgactc cggcacgttc 480  
ggcggctacc cggacatctg ccacgacaag tcctgggacc agtactggct ctgggcctcg 540  
caggagtccct acgcggccta cctgcgctcc atcggcatcg acgctggcg cttcgactac 600  
gtcaagggct acggggcctg ggtggtcaag gactggctca actgggtggg cggctggcg 660  
gtgggcgagt actgggacac caacgtcgac gcgctgctca actgggccta ctctccggc 720  
gccaaaggtgt tcgacttccc cctgtactac aagatggacg cggccttcga caacaagaac 780  
atcccggcgc tcgtcgaggc cctgaagaac ggcggcacgg tggctctccg cgacccttc 840  
aaggccgtga ccttcgtcgc caaccacgac acggacatca tctggaacaa gtaccggcg 900  
tacgccttca tcctcaccta cgagggccag cccacgatct tctaccgca ctacgaggag 960  
tggctgaaca aggacaagct caagaacctg atctggattc acgacaacct cgcgggcggc 1020  
tccactagta tcgtgtacta cgactccgac gagatgatct tcgtccgcaa cggctacggc 1080  
tccaagcccg gcctgatcac gtacatcaac ctgggctcct ccaaggtggg ccgctgggtg 1140  
tacgtcccga agttcgcccg cgcgtgcac caccagtaga ccggcaacct cggcggctgg 1200  
gtggacaagt acgtgtactc ctccggctgg gtctacctgg agggcccggc ctacgacccc 1260  
gccaacggcc agtacggcta ctccgtgtgg tcctactgcg gcgtcggc 1308

<210> 3  
<211> 800  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 3  
Met Gly His Trp Tyr Lys His Gln Arg Ala Tyr Gln Phe Thr Gly Glu  
1 5 10 15  
Asp Asp Phe Gly Lys Val Ala Val Val Lys Leu Pro Met Asp Leu Thr  
20 25 30  
Lys Val Gly Ile Ile Val Arg Leu Asn Glu Trp Gln Ala Lys Asp Val  
35 40 45  
Ala Lys Asp Arg Phe Ile Glu Ile Lys Asp Gly Lys Ala Glu Val Trp  
50 55 60  
Ile Leu Gln Gly Val Glu Glu Ile Phe Tyr Glu Lys Pro Asp Thr Ser

65					70					75				80	
Pro	Arg	Ile	Phe	Phe	Ala	Gln	Ala	Arg	Ser	Asn	Lys	Val	Ile	Glu	Ala
				85					90					95	
Phe	Leu	Thr	Asn	Pro	Val	Asp	Thr	Lys	Lys	Lys	Glu	Leu	Phe	Lys	Val
			100					105					110		
Thr	Val	Asp	Gly	Lys	Glu	Ile	Pro	Val	Ser	Arg	Val	Glu	Lys	Ala	Asp
		115					120					125			
Pro	Thr	Asp	Ile	Asp	Val	Thr	Asn	Tyr	Val	Arg	Ile	Val	Leu	Ser	Glu
		130				135					140				
Ser	Leu	Lys	Glu	Glu	Asp	Leu	Arg	Lys	Asp	Val	Glu	Leu	Ile	Ile	Glu
145					150					155					160
Gly	Tyr	Lys	Pro	Ala	Arg	Val	Ile	Met	Met	Glu	Ile	Leu	Asp	Asp	Tyr
			165						170					175	
Tyr	Tyr	Asp	Gly	Glu	Leu	Gly	Ala	Val	Tyr	Ser	Pro	Glu	Lys	Thr	Ile
			180				185						190		
Phe	Arg	Val	Trp	Ser	Pro	Val	Ser	Lys	Trp	Val	Lys	Val	Leu	Leu	Phe
		195					200					205			
Lys	Asn	Gly	Glu	Asp	Thr	Glu	Pro	Tyr	Gln	Val	Val	Asn	Met	Glu	Tyr
		210				215					220				
Lys	Gly	Asn	Gly	Val	Trp	Glu	Ala	Val	Val	Glu	Gly	Asp	Leu	Asp	Gly
225					230					235					240
Val	Phe	Tyr	Leu	Tyr	Gln	Leu	Glu	Asn	Tyr	Gly	Lys	Ile	Arg	Thr	Thr
			245						250					255	
Val	Asp	Pro	Tyr	Ser	Lys	Ala	Val	Tyr	Ala	Asn	Asn	Gln	Glu	Ser	Ala
		260						265					270		
Val	Val	Asn	Leu	Ala	Arg	Thr	Asn	Pro	Glu	Gly	Trp	Glu	Asn	Asp	Arg
		275					280					285			
Gly	Pro	Lys	Ile	Glu	Gly	Tyr	Glu	Asp	Ala	Ile	Ile	Tyr	Glu	Ile	His
		290				295					300				
Ile	Ala	Asp	Ile	Thr	Gly	Leu	Glu	Asn	Ser	Gly	Val	Lys	Asn	Lys	Gly
305					310					315					320
Leu	Tyr	Leu	Gly	Leu	Thr	Glu	Glu	Asn	Thr	Lys	Gly	Pro	Gly	Gly	Val
			325						330					335	
Thr	Thr	Gly	Leu	Ser	His	Leu	Val	Glu	Leu	Gly	Val	Thr	His	Val	His
			340					345					350		
Ile	Leu	Pro	Phe	Phe	Asp	Phe	Tyr	Thr	Gly	Asp	Glu	Leu	Asp	Lys	Asp
		355					360					365			
Phe	Glu	Lys	Tyr	Tyr	Asn	Trp	Gly	Tyr	Asp	Pro	Tyr	Leu	Phe	Met	Val
		370				375					380				
Pro	Glu	Gly	Arg	Tyr	Ser	Thr	Asp	Pro	Lys	Asn	Pro	His	Thr	Arg	Ile
385					390					395					400
Arg	Glu	Val	Lys	Glu	Met	Val	Lys	Ala	Leu	His	Lys	His	Gly	Ile	Gly
			405						410					415	
Val	Ile	Met	Asp	Met	Val	Phe	Pro	His	Thr	Tyr	Gly	Ile	Gly	Glu	Leu
		420					425					430			
Ser	Ala	Phe	Asp	Gln	Thr	Val	Pro	Tyr	Tyr	Phe	Tyr	Arg	Ile	Asp	Lys
		435					440					445			
Thr	Gly	Ala	Tyr	Leu	Asn	Glu	Ser	Gly	Cys	Gly	Asn	Val	Ile	Ala	Ser
		450				455					460				
Glu	Arg	Pro	Met	Met	Arg	Lys	Phe	Ile	Val	Asp	Thr	Val	Thr	Tyr	Trp
465					470					475					480
Val	Lys	Glu	Tyr	His	Ile	Asp	Gly	Phe	Arg	Phe	Asp	Gln	Met	Gly	Leu
			485						490					495	
Ile	Asp	Lys	Lys	Thr	Met	Leu	Glu	Val	Glu	Arg	Ala	Leu	His	Lys	Ile

500										505										510																											
Asp	Pro	Thr	Ile	Ile	Leu	Tyr	Gly	Glu	Pro	Trp	Gly	Gly	Trp	Gly	Ala	Asp	Pro	Thr	Ile	Ile	Leu	Tyr	Gly	Glu	Pro	Trp	Gly	Gly	Trp	Gly	Ala	Asp	Pro	Thr	Ile	Ile	Leu	Tyr	Gly	Glu	Pro	Trp	Gly	Gly	Trp	Gly	Ala
515										520										525																											
Pro	Ile	Arg	Phe	Gly	Lys	Ser	Asp	Val	Ala	Gly	Thr	His	Val	Ala	Ala	Pro	Ile	Arg	Phe	Gly	Lys	Ser	Asp	Val	Ala	Gly	Thr	His	Val	Ala	Ala	Pro	Ile	Arg	Phe	Gly	Lys	Ser	Asp	Val	Ala	Gly	Thr	His	Val	Ala	Ala
530										535										540																											
Phe	Asn	Asp	Glu	Phe	Arg	Asp	Ala	Ile	Arg	Gly	Ser	Val	Phe	Asn	Pro	Phe	Asn	Asp	Glu	Phe	Arg	Asp	Ala	Ile	Arg	Gly	Ser	Val	Phe	Asn	Pro	Phe	Asn	Asp	Glu	Phe	Arg	Asp	Ala	Ile	Arg	Gly	Ser	Val	Phe	Asn	Pro
545										550										555																											
Ser	Val	Lys	Gly	Phe	Val	Met	Gly	Gly	Tyr	Gly	Lys	Glu	Thr	Lys	Ile	Ser	Val	Lys	Gly	Phe	Val	Met	Gly	Gly	Tyr	Gly	Lys	Glu	Thr	Lys	Ile	Ser	Val	Lys	Gly	Phe	Val	Met	Gly	Gly	Tyr	Gly	Lys	Glu	Thr	Lys	Ile
565										570										575																											
Lys	Arg	Gly	Val	Val	Gly	Ser	Ile	Asn	Tyr	Asp	Gly	Lys	Leu	Ile	Lys	Lys	Arg	Gly	Val	Val	Gly	Ser	Ile	Asn	Tyr	Asp	Gly	Lys	Leu	Ile	Lys	Lys	Arg	Gly	Val	Val	Gly	Ser	Ile	Asn	Tyr	Asp	Gly	Lys	Leu	Ile	Lys
580										585										590																											
Ser	Phe	Ala	Leu	Asp	Pro	Glu	Glu	Thr	Ile	Asn	Tyr	Ala	Ala	Cys	His	Ser	Phe	Ala	Leu	Asp	Pro	Glu	Glu	Thr	Ile	Asn	Tyr	Ala	Ala	Cys	His	Ser	Phe	Ala	Leu	Asp	Pro	Glu	Glu	Thr	Ile	Asn	Tyr	Ala	Ala	Cys	His
595										600										605																											
Asp	Asn	His	Thr	Leu	Trp	Asp	Lys	Asn	Tyr	Leu	Ala	Ala	Lys	Ala	Asp	Asp	Asn	His	Thr	Leu	Trp	Asp	Lys	Asn	Tyr	Leu	Ala	Ala	Lys	Ala	Asp	Asp	Asn	His	Thr	Leu	Trp	Asp	Lys	Asn	Tyr	Leu	Ala	Ala	Lys	Ala	Asp
610										615										620																											
Lys	Lys	Lys	Glu	Trp	Thr	Glu	Glu	Glu	Leu	Lys	Asn	Ala	Gln	Lys	Leu	Lys	Lys	Lys	Glu	Trp	Thr	Glu	Glu	Glu	Leu	Lys	Asn	Ala	Gln	Lys	Leu	Lys	Lys	Lys	Glu	Trp	Thr	Glu	Glu	Glu	Leu	Lys	Asn	Ala	Gln	Lys	Leu
625										630										635																											
Ala	Gly	Ala	Ile	Leu	Leu	Thr	Ser	Gln	Gly	Val	Pro	Phe	Leu	His	Gly	Ala	Gly	Ala	Ile	Leu	Leu	Thr	Ser	Gln	Gly	Val	Pro	Phe	Leu	His	Gly	Ala	Gly	Ala	Ile	Leu	Leu	Thr	Ser	Gln	Gly	Val	Pro	Phe	Leu	His	Gly
645										650										655																											
Gly	Gln	Asp	Phe	Cys	Arg	Thr	Thr	Asn	Phe	Asn	Asp	Asn	Ser	Tyr	Asn	Gly	Gln	Asp	Phe	Cys	Arg	Thr	Thr	Asn	Phe	Asn	Asp	Asn	Ser	Tyr	Asn	Gly	Gln	Asp	Phe	Cys	Arg	Thr	Thr	Asn	Phe	Asn	Asp	Asn	Ser	Tyr	Asn
660										665										670																											
Ala	Pro	Ile	Ser	Ile	Asn	Gly	Phe	Asp	Tyr	Glu	Arg	Lys	Leu	Gln	Phe	Ala	Pro	Ile	Ser	Ile	Asn	Gly	Phe	Asp	Tyr	Glu	Arg	Lys	Leu	Gln	Phe	Ala	Pro	Ile	Ser	Ile	Asn	Gly	Phe	Asp	Tyr	Glu	Arg	Lys	Leu	Gln	Phe
675										680										685																											
Ile	Asp	Val	Phe	Asn	Tyr	His	Lys	Gly	Leu	Ile	Lys	Leu	Arg	Lys	Glu	Ile	Asp	Val	Phe	Asn	Tyr	His	Lys	Gly	Leu	Ile	Lys	Leu	Arg	Lys	Glu	Ile	Asp	Val	Phe	Asn	Tyr	His	Lys	Gly	Leu	Ile	Lys	Leu	Arg	Lys	Glu
690										695										700																											
His	Pro	Ala	Phe	Arg	Leu	Lys	Asn	Ala	Glu	Glu	Ile	Lys	Lys	His	Leu	His	Pro	Ala	Phe	Arg	Leu	Lys	Asn	Ala	Glu	Glu	Ile	Lys	Lys	His	Leu	His	Pro	Ala	Phe	Arg	Leu	Lys	Asn	Ala	Glu	Glu	Ile	Lys	Lys	His	Leu
705										710										715																											
Glu	Phe	Leu	Pro	Gly	Arg	Arg	Ile	Val	Ala	Phe	Met	Leu	Lys	Asp	Glu	Phe	Leu	Pro	Gly	Arg	Arg	Ile	Val	Ala	Phe	Met	Leu	Lys	Asp	Glu	Phe	Leu	Pro	Gly	Arg	Arg	Ile	Val	Ala	Phe	Met	Leu	Lys	Asp			
725										730										735																											
His	Ala	Gly	Gly	Asp	Pro	Trp	Lys	Asp	Ile	Val	Val	Ile	Tyr	Asn	Gly	His	Ala	Gly	Gly	Asp	Pro	Trp	Lys	Asp	Ile	Val	Val	Ile	Tyr	Asn	Gly	His	Ala	Gly	Gly	Asp	Pro	Trp	Lys	Asp	Ile	Val	Val	Ile	Tyr	Asn	Gly
740										745										750																											
Asn	Leu	Glu	Lys	Thr	Thr	Tyr	Lys	Leu	Pro	Glu	Gly	Lys	Trp	Asn	Val	Asn	Leu	Glu	Lys	Thr	Thr	Tyr	Lys	Leu	Pro	Glu	Gly	Lys	Trp	Asn	Val	Asn	Leu	Glu	Lys	Thr	Thr	Tyr	Lys	Leu	Pro	Glu	Gly	Lys	Trp	Asn	Val
755										760										765																											
Val	Val	Asn	Ser	Gln	Lys	Ala	Gly	Thr	Glu	Val	Ile	Glu	Thr	Val	Glu	Val	Val	Asn	Ser	Gln	Lys	Ala	Gly	Thr	Glu	Val	Ile	Glu	Thr	Val	Glu	Val	Val	Asn	Ser	Gln	Lys	Ala	Gly	Thr	Glu	Val	Ile	Glu	Thr	Val	Glu
770										775										780																											
Gly	Thr	Ile	Glu	Leu	Asp	Pro	Leu	Ser	Ala	Tyr	Val	Leu	Tyr	Arg	Glu	Gly	Thr	Ile	Glu	Leu	Asp	Pro	Leu	Ser	Ala	Tyr	Val	Leu	Tyr	Arg	Glu	Gly	Thr	Ile	Glu	Leu	Asp	Pro	Leu	Ser	Ala	Tyr	Val	Leu	Tyr	Arg	Glu
785										790										795																											

&lt;210&gt; 4

&lt;211&gt; 2400

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic

&lt;400&gt; 4

```

atgggccact ggtacaagca ccagcgcgcc taccagttca ccggcgagga cgacttcggg 60
aagggtggcg tggatgaagct cccgatggac ctcaccaagg tgggcatcat cgtgcgcctc 120
aacgagtggc aggcgaagga cgtggccaag gaccgcttca tcgagatcaa ggacggcaag 180
gccgaggtgt ggatactcca gggcgtggag gagatcttct acgagaagcc ggacacctcc 240
ccgcgcacat tcttcgcccc ggcccgtccc aacaagggtga tcgaggcctt cctcaccaac 300
ccggtggaca ccaagaagaa ggagctgttc aagggtgaccg tcgacggcaa ggagatcccc 360

```

```

gtgtcccgcg tggagaaggg cgacccgacc gacatcgacg tgaccaacta cgtgcgcac 420
gtgctctccg agtccctcaa ggaggaggac ctccgcaagg acgtggagct gatcatcgag 480
ggctacaagc cggcccgcgt gatcatgatg gagatcctcg acgactacta ctacgacggc 540
gagctggggg cgggtgtactc cccggagaag accatcttcc gcgtgtggtc cccgggtgtcc 600
aagtgggtga aggtgtctct cttcaagaac ggcgaggaca ccgagccgta ccaggtggtg 660
aacatggagt acaaggggcaa cggcgtgtgg gaggccgtgg tggagggcga cctcgacggc 720
gtgttctacc tctaccagct ggagaactac ggcaagatcc gcaccaccgt ggaccctac 780
tccaaggccg tgtacgccaa caaccaggag tctgcagtgg tgaacctcgc ccgcaccaac 840
ccggaggggt gggagaacga ccgcgcccg aagatcgagg gctacgagga cgccatcatc 900
tacgagatcc acatcgccga catcaccggc ctggagaact ccggcgtgaa gaacaagggc 960
ctctacctcg gcctcaccga ggagaacacc aaggccccgg gcggcgtgac caccggcctc 1020
tcccacctcg tggagctggg cgtgaccacac gtgcacatcc tcccgttctt cgacttctac 1080
accggcgacg agctggacaa ggacttctgag aagtactaca actggggcta cgaccctac 1140
ctcttcatgg tgcggagggg ccgctactcc accgaccgga agaaccgcga caccggaatt 1200
cgcgaggtga aggagatggg gaaggccctc cacaagcacg gcatcggcgt gatcatggac 1260
atggtgttcc cgcacacctc cggcatcggc gagctgtccg ccttcgacca gaccgtgccg 1320
tactacttct accgcatcga caagaccggc gcctacctca acgagtcgg ctgcggcaac 1380
gtgatcgcc cggagcgccc gatgatgcgc aagttcatcg tggacaccgt gacctactgg 1440
gtgaaggagt accacatcga cggcttccgc ttcgaccaga tgggcctcat cgacaagaag 1500
accatgctgg aggtggagcg cgccctccac aagatcgacc cgaccatcat cctctacggc 1560
gagccgtggg gcggctgggg ggccccgatc cgcttcggca agtccgacgt ggccggcacc 1620
cacgtggccg cttcaacga cgagtccgc gacgccatcc gcggctccgt gttcaaccg 1680
tccgtgaagg gcttcgtgat gggcggtac ggcaaggaga ccaagatcaa gcgcggcg 1740
gtgggctcca tcaactacga cggcaagctc atcaagtcct tcgccctcga cccggaggag 1800
accatcaact acgcccctg ccacgacaac cacaccctct gggacaagaa ctacctcgcc 1860
gccaaaggcc acaagaagaa ggagtggacc gaggaggagc tgaagaacgc ccagaagctc 1920
gccgcgccca tctcctcac tagtcagggc gtgcccgttc tccacggcgg ccaggacttc 1980
tgccgcacca ccaacttcaa cgacaactcc tacaacgccc cgatctccat caacggcttc 2040
gactacgagc gcaagctcca gttcatcgac gtgttcaact accacaagg cctcatcaag 2100
ctccgcaagg agcaccggc cttccgcctc aagaacggc aggagatcaa gaagcacctg 2160
gagttcctcc cggcgggcg ccgcatcgtg gccttcatgc tcaaggacca cgccggcgcc 2220
gaccctgga aggacatcgt ggtgatctac aacggcaacc tggagaagac cacctacaag 2280
ctcccggagg gcaagtggaa cgtggtgtg aactcccaga aggccggcac cgaggtgatc 2340
gagaccgtgg agggcaccat cgagctggac ccgctctccg cctacgtgct ctaccgcgag 2400

```

&lt;210&gt; 5

&lt;211&gt; 693

&lt;212&gt; PRT

<213> *Sulfolobus solfataricus*

&lt;400&gt; 5

```

Met Glu Thr Ile Lys Ile Tyr Glu Asn Lys Gly Val Tyr Lys Val Val
 1           5           10          15
Ile Gly Glu Pro Phe Pro Pro Ile Glu Phe Pro Leu Glu Gln Lys Ile
      20           25           30
Ser Ser Asn Lys Ser Leu Ser Glu Leu Gly Leu Thr Ile Val Gln Gln
      35           40           45
Gly Asn Lys Val Ile Val Glu Lys Ser Leu Asp Leu Lys Glu His Ile
      50           55           60
Ile Gly Leu Gly Glu Lys Ala Phe Glu Leu Asp Arg Lys Arg Lys Arg
      65           70           75           80
Tyr Val Met Tyr Asn Val Asp Ala Gly Ala Tyr Lys Lys Tyr Gln Asp
      85           90           95

```



Pro Leu Tyr Val Ser Ile Pro Leu Phe Ile Ser Val Lys Asp Gly Val  
                   100                  105                  110  
 Ala Thr Gly Tyr Phe Phe Asn Ser Ala Ser Lys Val Ile Phe Asp Val  
                   115                  120                  125  
 Gly Leu Glu Glu Tyr Asp Lys Val Ile Val Thr Ile Pro Glu Asp Ser  
                   130                  135                  140  
 Val Glu Phe Tyr Val Ile Glu Gly Pro Arg Ile Glu Asp Val Leu Glu  
 145                  150                  155                  160  
 Lys Tyr Thr Glu Leu Thr Gly Lys Pro Phe Leu Pro Pro Met Trp Ala  
                   165                  170                  175  
 Phe Gly Tyr Met Ile Ser Arg Tyr Ser Tyr Tyr Pro Gln Asp Lys Val  
                   180                  185                  190  
 Val Glu Leu Val Asp Ile Met Gln Lys Glu Gly Phe Arg Val Ala Gly  
                   195                  200                  205  
 Val Phe Leu Asp Ile His Tyr Met Asp Ser Tyr Lys Leu Phe Thr Trp  
 210                  215                  220  
 His Pro Tyr Arg Phe Pro Glu Pro Lys Lys Leu Ile Asp Glu Leu His  
 225                  230                  235                  240  
 Lys Arg Asn Val Lys Leu Ile Thr Ile Val Asp His Gly Ile Arg Val  
                   245                  250                  255  
 Asp Gln Asn Tyr Ser Pro Phe Leu Ser Gly Met Gly Lys Phe Cys Glu  
                   260                  265                  270  
 Ile Glu Ser Gly Glu Leu Phe Val Gly Lys Met Trp Pro Gly Thr Thr  
                   275                  280                  285  
 Val Tyr Pro Asp Phe Phe Arg Glu Asp Thr Arg Glu Trp Trp Ala Gly  
                   290                  295                  300  
 Leu Ile Ser Glu Trp Leu Ser Gln Gly Val Asp Gly Ile Trp Leu Asp  
 305                  310                  315                  320  
 Met Asn Glu Pro Thr Asp Phe Ser Arg Ala Ile Glu Ile Arg Asp Val  
                   325                  330                  335  
 Leu Ser Ser Leu Pro Val Gln Phe Arg Asp Asp Arg Leu Val Thr Thr  
                   340                  345                  350  
 Phe Pro Asp Asn Val Val His Tyr Leu Arg Gly Lys Arg Val Lys His  
                   355                  360                  365  
 Glu Lys Val Arg Asn Ala Tyr Pro Leu Tyr Glu Ala Met Ala Thr Phe  
                   370                  375                  380  
 Lys Gly Phe Arg Thr Ser His Arg Asn Glu Ile Phe Ile Leu Ser Arg  
 385                  390                  395                  400  
 Ala Gly Tyr Ala Gly Ile Gln Arg Tyr Ala Phe Ile Trp Thr Gly Asp  
                   405                  410                  415  
 Asn Thr Pro Ser Trp Asp Asp Leu Lys Leu Gln Leu Gln Leu Val Leu  
                   420                  425                  430  
 Gly Leu Ser Ile Ser Gly Val Pro Phe Val Gly Cys Asp Ile Gly Gly  
                   435                  440                  445  
 Phe Gln Gly Arg Asn Phe Ala Glu Ile Asp Asn Ser Met Asp Leu Leu  
                   450                  455                  460  
 Val Lys Tyr Tyr Ala Leu Ala Leu Phe Phe Pro Phe Tyr Arg Ser His  
 465                  470                  475                  480  
 Lys Ala Thr Asp Gly Ile Asp Thr Glu Pro Val Phe Leu Pro Asp Tyr  
                   485                  490                  495  
 Tyr Lys Glu Lys Val Lys Glu Ile Val Glu Leu Arg Tyr Lys Phe Leu  
                   500                  505                  510  
 Pro Tyr Ile Tyr Ser Leu Ala Leu Glu Ala Ser Glu Lys Gly His Pro  
                   515                  520                  525

Val Ile Arg Pro Leu Phe Tyr Glu Phe Gln Asp Asp Asp Asp Met Tyr  
 530 535 540  
 Arg Ile Glu Asp Glu Tyr Met Val Gly Lys Tyr Leu Leu Tyr Ala Pro  
 545 550 555 560  
 Ile Val Ser Lys Glu Glu Ser Arg Leu Val Thr Leu Pro Arg Gly Lys  
 565 570 575  
 Trp Tyr Asn Tyr Trp Asn Gly Glu Ile Ile Asn Gly Lys Ser Val Val  
 580 585 590  
 Lys Ser Thr His Glu Leu Pro Ile Tyr Leu Arg Glu Gly Ser Ile Ile  
 595 600 605  
 Pro Leu Glu Gly Asp Glu Leu Ile Val Tyr Gly Glu Thr Ser Phe Lys  
 610 615 620  
 Arg Tyr Asp Asn Ala Glu Ile Thr Ser Ser Ser Asn Glu Ile Lys Phe  
 625 630 635 640  
 Ser Arg Glu Ile Tyr Val Ser Lys Leu Thr Ile Thr Ser Glu Lys Pro  
 645 650 655  
 Val Ser Lys Ile Ile Val Asp Asp Ser Lys Glu Ile Gln Val Glu Lys  
 660 665 670  
 Thr Met Gln Asn Thr Tyr Val Ala Lys Ile Asn Gln Lys Ile Arg Gly  
 675 680 685  
 Lys Ile Asn Leu Glu  
 690

&lt;210&gt; 6

&lt;211&gt; 2082

&lt;212&gt; DNA

<213> *Sulfolobus solfataricus*

&lt;400&gt; 6

atggagacca tcaagatcta cgagaacaag ggctgtgtaca aggtgggtgat cggcgagccg 60  
 ttcccgccga tcgagttccc gctcgagcag aagatctcct ccaacaagtc cctctccgag 120  
 ctgggcctca ccacgtgca gcagggaac aaggtgatcg tggagaagtc cctcgacctc 180  
 aaggagcaca tcacggcct cggcgagaag gccttcgagc tggaccgcaa gcgcaagcgc 240  
 tacgtgatgt acaacgtgga cgccggcgcc tacaagaagt accaggaccc gctctacgtg 300  
 tccatccgcg tcttcatttc cgtgaaggac ggctgtggca ccggctactt cttcaactcc 360  
 gcctccaagg tgatcttcga cgtgggcctc gaggagtacg acaaggtgat cgtgacctac 420  
 ccggaggact ccgtggagtt ctacgtgatc gagggcccgc gcatcgagga cgtgctcgag 480  
 aagtacaccg agctgaccgg caagccgttc ctcccgcga tgtgggcctt cggctacatg 540  
 atctcccgct actcctacta cccgcaggac aaggtgggtg agctgggtga catcatgcag 600  
 aaggagggct tccgcgtggc cggcgtgttc ctcgacatcc actacatgga ctccctacaag 660  
 ctcttcacct ggcacccgta ccgcttcccg gagccgaaga agctcatcga cgagctgcac 720  
 aagcgcaacg tgaagctcat caccatcggt gaccacggca tccgcgtgga ccagaactac 780  
 tccccgttcc tctccggcat gggcaagtcc tgcgagatcg agtccggcga gctgttcgtg 840  
 ggcaagatgt ggccgggcac caccgtgtac ccggacttct tccgcgagga caccgcgag 900  
 tgggtgggccc gcctcatctc cgagtggctc tcccagggcg tggacggcat ctggctcgac 960  
 atgaacgagc cgaccgactt ctcccgcgcc atcgagatcc gcgacgtgct ctccctccctc 1020  
 ccggtgcagt tccgcgacga ccgcctcggt accaccttcc cggacaacgt ggtgcactac 1080  
 ctccgcggca agcgcgtgaa gcacgagaag gtgcgcaacg cctaccgcgt ctacgaggcg 1140  
 atggccacct tcaagggtct ccgcacctcc caccgcaacg agatcttcat cctctcccgc 1200  
 gccggctacg ccggcatcca gcgctacgcc ttcattctgga ccggcgacaa caccctcgcc 1260  
 tgggacgacc tcaagctcca gctccagctc gtgctcggcc tctccatctc cggcgtgccg 1320  
 ttctgtgggt gcgacatcgg cggcttccag gcccgcaact tcgccgagat cgacaactcg 1380  
 atggacctcc tcgtgaagta ctacgccctc gccctcttct tcccgttcta ccgctccac 1440

```

aagggccaccg acggcatcga caccgagccg gtgttcctcc cggactacta caaggagaag 1500
gtgaaggaga tcgtggagct gcgctacaag ttctctcccg acatctactc cctcgccctc 1560
gaggcctccg agaagggcca cccgggtgatc cgcccgtctt tctacgagtt ccaggacgac 1620
gacgacatgt accgcatcga ggacgagtag atggtgggca agtacctcct ctacgccccg 1680
atcgtgtcca aggaggagtc cgcctcgtg accctccgc gcggcaagtg gtacaactac 1740
tggaacggcg agatcatcaa cggcaagtcc gtggtgaagt ccacccacga gctgccgac 1800
tacctccgcg agggctccat catcccgctc gagggcgacg agctgacgtg gtacggcgag 1860
acctcttca agcgctacga caacgccgag atcacctcct cctccaacga gatcaagttc 1920
tcccgcgaga tctacgtgtc caagctcacc atcacctccg agaagccggt gtccaagatc 1980
atcgtggacg actccaagga gatccaggtg gagaagacca tgcagaacac ctacgtggcc 2040
aagatcaacc agaagatccg cggcaagatc aacctcgagt ga 2082

```

&lt;210&gt; 7

&lt;211&gt; 1818

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic

&lt;400&gt; 7

```

atggcggtc tggccacgtc gcagctcgtc gcaacgcgcg ccggcctggg cgtcccggac 60
gcgtccacgt tccgcgcg cgcccgagc ggcctgaggg gggcccgggc gtcggcgcg 120
gcggacacgc tcagcatgcg gaccagcgcg cgcgcggcgc ccaggcacca gcaccagcag 180
gcgcgcgcgc gggccaggtt cccgtcgtc gtctgtgtcg ccagcgcgcg catgaacgtc 240
gtcttcgtcg gcgcgagat ggcgcgtgg agcaagaccg gaggcctcgg cgacgtcctc 300
ggcgcctgc cgcgcggcat ggcgcggaac gggcaccgtg tcatggctgt ctctccccgc 360
tacgaccagt acaaggacgc ctgggacacc agcgtcgtgt ccgagatcaa gatgggagac 420
gggtacgaga cggtcagggt ctccactgc tacaagcgcg gagtggaccg cgtgttcgtt 480
gaccacccac tgttcttggg gaggggttgg ggaagaccg aggagaagat ctacgggcct 540
gtcgtggaa cggactacag ggacaaccag ctgcggttca gcctgctatg ccaggcagca 600
cttgaagctc caaggatcct gagcctcaac aacaacccat actctcctcg accatacggg 660
gaggacgtcg tgttcgtctg caacgactgg cacaccggcc ctctctcgtg ctacctcaag 720
agcaactacc agtcccacgg catctacagg gacgcaaaga ccgctttctg catccacaac 780
atctcctacc agggcgggtt cgccttctcc gactaccgg agctgaacct ccccgagaga 840
ttcaagtcgt ccttcgattt catcgacggc tacgagaagc ccgtggaagg ccggaagatc 900
aactggatga aggcgggat cctcgaggcc gacaggtcc tcaccgtcag cccctactac 960
gccgaggagc tcatctccgg catcgccagg ggctgcgagc tcgacaacat catgcgcctc 1020
accggcatca ccggcatcgt caacggcatg gacgtcagcg agtgggaccc cagcaggggac 1080
aagtacatcg ccgtgaagta cgacgtgtcg acggccgtgg aggccaaggc gctgaacaag 1140
gaggcgtgc aggcggaggt cgggctcccg gtggaccgga acatcccgtt ggtggcggtt 1200
atcggcaggc tggaaagagca gaagggcccc gacgtcatgg cggccgcat cccgcagctc 1260
atggagatgg tggaggacgt gcagatcgtt ctgctgggca cgggcaagaa gaagttcgag 1320
cgcatgctca tgagcgccga ggagaagtcc ccaggcaagg tgcgcgccgt ggtcaagtcc 1380
aacgcggcgc tggcgacca catcatggcc ggcgcggacg tgctcgccgt caccagccgc 1440
ttcgagccct gcggcctcat ccagctgcag gggatgcgat acggaacgcc ctgcgcctgc 1500
gcgtccaccg gtggactcgt cgacaccatc atcgaaggca agaccgggtt ccacatggg 1560
cgctcagcg tcgactgcaa cgtcgtggag ccggcgagc tcaagaaggt ggccaccacc 1620
ttgcagcgcg ccatcaaggt ggtcggcacg ccggcgtagc aggagatggt gaggaactgc 1680
atgatccagg atctctcctg gaaggccct gccaaagaact gggagaacgt gctgctcagc 1740
ctcggggtcg ccggcggcga gccagggggt gaaggcgagg agatcgcgcc gctcgccaag 1800
gagaacgtgg ccgcgcc 1818

```

&lt;210&gt; 8

<211> 606  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 8

```

Met Ala Ala Leu Ala Thr Ser Gln Leu Val Ala Thr Arg Ala Gly Leu
 1          5          10          15
Gly Val Pro Asp Ala Ser Thr Phe Arg Arg Gly Ala Ala Gln Gly Leu
          20          25          30
Arg Gly Ala Arg Ala Ser Ala Ala Ala Asp Thr Leu Ser Met Arg Thr
          35          40          45
Ser Ala Arg Ala Ala Pro Arg His Gln His Gln Gln Ala Arg Arg Gly
 50          55          60
Ala Arg Phe Pro Ser Leu Val Val Cys Ala Ser Ala Gly Met Asn Val
 65          70          75          80
Val Phe Val Gly Ala Glu Met Ala Pro Trp Ser Lys Thr Gly Gly Leu
          85          90          95
Gly Asp Val Leu Gly Gly Leu Pro Pro Ala Met Ala Ala Asn Gly His
          100          105          110
Arg Val Met Val Val Ser Pro Arg Tyr Asp Gln Tyr Lys Asp Ala Trp
          115          120          125
Asp Thr Ser Val Val Ser Glu Ile Lys Met Gly Asp Gly Tyr Glu Thr
          130          135          140
Val Arg Phe Phe His Cys Tyr Lys Arg Gly Val Asp Arg Val Phe Val
 145          150          155          160
Asp His Pro Leu Phe Leu Glu Arg Val Trp Gly Lys Thr Glu Glu Lys
          165          170          175
Ile Tyr Gly Pro Val Ala Gly Thr Asp Tyr Arg Asp Asn Gln Leu Arg
          180          185          190
Phe Ser Leu Leu Cys Gln Ala Ala Leu Glu Ala Pro Arg Ile Leu Ser
          195          200          205
Leu Asn Asn Asn Pro Tyr Phe Ser Gly Pro Tyr Gly Glu Asp Val Val
          210          215          220
Phe Val Cys Asn Asp Trp His Thr Gly Pro Leu Ser Cys Tyr Leu Lys
 225          230          235          240
Ser Asn Tyr Gln Ser His Gly Ile Tyr Arg Asp Ala Lys Thr Ala Phe
          245          250          255
Cys Ile His Asn Ile Ser Tyr Gln Gly Arg Phe Ala Phe Ser Asp Tyr
          260          265          270
Pro Glu Leu Asn Leu Pro Glu Arg Phe Lys Ser Ser Phe Asp Phe Ile
          275          280          285
Asp Gly Tyr Glu Lys Pro Val Glu Gly Arg Lys Ile Asn Trp Met Lys
          290          295          300
Ala Gly Ile Leu Glu Ala Asp Arg Val Leu Thr Val Ser Pro Tyr Tyr
 305          310          315          320
Ala Glu Glu Leu Ile Ser Gly Ile Ala Arg Gly Cys Glu Leu Asp Asn
          325          330          335
Ile Met Arg Leu Thr Gly Ile Thr Gly Ile Val Asn Gly Met Asp Val
          340          345          350
Ser Glu Trp Asp Pro Ser Arg Asp Lys Tyr Ile Ala Val Lys Tyr Asp
          355          360          365

```

Val Ser Thr Ala Val Glu Ala Lys Ala Leu Asn Lys Glu Ala Leu Gln  
 370 375 380  
 Ala Glu Val Gly Leu Pro Val Asp Arg Asn Ile Pro Leu Val Ala Phe  
 385 390 395 400  
 Ile Gly Arg Leu Glu Glu Gln Lys Gly Pro Asp Val Met Ala Ala Ala  
 405 410 415  
 Ile Pro Gln Leu Met Glu Met Val Glu Asp Val Gln Ile Val Leu Leu  
 420 425 430  
 Gly Thr Gly Lys Lys Lys Phe Glu Arg Met Leu Met Ser Ala Glu Glu  
 435 440 445  
 Lys Phe Pro Gly Lys Val Arg Ala Val Val Lys Phe Asn Ala Ala Leu  
 450 455 460  
 Ala His His Ile Met Ala Gly Ala Asp Val Leu Ala Val Thr Ser Arg  
 465 470 475 480  
 Phe Glu Pro Cys Gly Leu Ile Gln Leu Gln Gly Met Arg Tyr Gly Thr  
 485 490 495  
 Pro Cys Ala Cys Ala Ser Thr Gly Gly Leu Val Asp Thr Ile Ile Glu  
 500 505 510  
 Gly Lys Thr Gly Phe His Met Gly Arg Leu Ser Val Asp Cys Asn Val  
 515 520 525  
 Val Glu Pro Ala Asp Val Lys Lys Val Ala Thr Thr Leu Gln Arg Ala  
 530 535 540  
 Ile Lys Val Val Gly Thr Pro Ala Tyr Glu Glu Met Val Arg Asn Cys  
 545 550 555 560  
 Met Ile Gln Asp Leu Ser Trp Lys Gly Pro Ala Lys Asn Trp Glu Asn  
 565 570 575  
 Val Leu Leu Ser Leu Gly Val Ala Gly Gly Glu Pro Gly Val Glu Gly  
 580 585 590  
 Glu Glu Ile Ala Pro Leu Ala Lys Glu Asn Val Ala Ala Pro  
 595 600 605

&lt;210&gt; 9

&lt;211&gt; 2223

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic

&lt;400&gt; 9

atggccaagt acctggagct ggaggagggc ggcgatgatca tgcaggcggt ctactgggac 60  
 gtcccgagcg gaggcacatct gtgggacacc atccgccaga agatccccga gtggtacgac 120  
 gccggcatct ccgcgatctg gataccgccca gcttccaagg gcatgtccgg gggctactcg 180  
 atgggctacg acccgtagca ctacttcgac ctccggcgagt actaccagaa gggcacgggtg 240  
 gagacgcgct tcgggtccaa gcaggagctc atcaacatga tcaacacggc gcacgcctac 300  
 ggcatcaagg tcatcgcgga catcgtgatc aaccacaggg ccggcgggcga cctggagtg 360  
 aaccggttcg tcggcgacta cacctggacg gacttctcca aggtcgctc cggaagtac 420  
 accgccaact acctcgactt ccacccaac gagctgcacg cgggcgactc cggcacgttc 480  
 ggcggtacc cggacatctg ccacgacaag tcctgggacc agtactgggt ctgggcctcg 540  
 caggagtcct acgcggccta cctgcgtctc atcggcatcg acgcgtggcg ctctgactac 600  
 gtcaagggtc acggggcctg ggtggtcaag gactggctca actggtgggg cggctggggc 660  
 gtgggcgagt actgggacac caacgtcgac gcgtgctca actgggccta ctctccggc 720  
 gccaaagtgt tcgacttccc cctgtactac aagatggacg cggccttcga caacaagaac 780

```

atccccggcgc tcgtcgaggg cctgaagaac ggcggcacgg tgggtctccc cgacccgttc 840
aaggccgtga ccttcgtcgc caaccacgac acggacatca tctggaacaa gtacccggcg 900
tacgccttca tcctcaccta cgagggccag cccacgatct tctaccgca ctacgaggag 960
tggctgaaca aggacaagct caagaacctg atctggattc acgacaacct cgcggggcgc 1020
tccactagta tcgtgtacta cgactccgac gagatgatct tcgtccgcaa cggctacggc 1080
tccaagcccc gcctgatcac gtacatcaac ctgggctcct ccaaggtggg ccgctgggtg 1140
tacgtcccga agttcgccgg cgcgtgcac cagagtaca ccggcaacct cggcggctgg 1200
gtggacaagt acgtgtactc ctccggtgg gtctacctgg agggcccgcc ctacgacccc 1260
gccaacggcc agtacggcta ctccgtgtgg tcctactgcg gcgtcggcac atcgattgct 1320
ggcatcctcg agggcgacag ggtcctcacc gtcagccctt actacgccga ggagctcatc 1380
tccggcatcg ccaggggctg cgagctcgac aacatcatgc gcctcaccgg catcaccggc 1440
atcgtaacg gcattggacgt cagcgagtgg gacc'ccagca gggacaagta catcgccgtg 1500
aagtacgacg tgtcgacggc cgtggaggcc aaggcgctga acaaggaggc gctgcaggcg 1560
gaggtcgggc tcccgtgga ccggaacatc ccgctgggtg cgttcacgag caggctggaa 1620
gagcagaagg gccccgacgt catggcggcc gccatcccgc agtcatgga gatggtggag 1680
gacgtgcaga tcgttctgct gggcacgggc aagaagaagt tcgagcgcat gctcatgagc 1740
gccgaggaga agtcccagg caaggtgctg gccgtggtca agttcaacgc ggcgctggcg 1800
caccacatca tggccggcgc cgacgtgctc gccgtcacca gccgcttcga gccctgccc 1860
ctcatccagc tgcaggggat gcgatacgga acgccctgcg cctgcgcgtc caccggtgga 1920
ctcgtcgaca ccacatcgga aggcaagacc ggggttcaca tgggccgcct cagcgtcgac 1980
tgcaacgctg tggagccggc ggacgtcaag aagggtggca ccaccttgca gcgcgccatc 2040
aagggtgctg gcacgccggc gtacgaggag atggtgagga actgcatgat ccaggatctc 2100
tcctggaagg gccctgccaa gaactgggag aacgtgctgc tcagcctcgg ggtcgccggc 2160
ggcgagccag ggggtgaagg cgaggagatc gcgccgctcg ccaaggagaa cgtggccgcg 2220
ccc
2223

```

&lt;210&gt; 10

&lt;211&gt; 741

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic

&lt;400&gt; 10

```

Met Ala Lys Tyr Leu Glu Leu Glu Gly Gly Val Ile Met Gln Ala
1           5           10           15
Phe Tyr Trp Asp Val Pro Ser Gly Gly Ile Trp Trp Asp Thr Ile Arg
20           25           30
Gln Lys Ile Pro Glu Trp Tyr Asp Ala Gly Ile Ser Ala Ile Trp Ile
35           40           45
Pro Pro Ala Ser Lys Gly Met Ser Gly Gly Tyr Ser Met Gly Tyr Asp
50           55           60
Pro Tyr Asp Tyr Phe Asp Leu Gly Glu Tyr Tyr Gln Lys Gly Thr Val
65           70           75           80
Glu Thr Arg Phe Gly Ser Lys Gln Glu Leu Ile Asn Met Ile Asn Thr
85           90           95
Ala His Ala Tyr Gly Ile Lys Val Ile Ala Asp Ile Val Ile Asn His
100          105          110
Arg Ala Gly Gly Asp Leu Glu Trp Asn Pro Phe Val Gly Asp Tyr Thr
115          120          125
Trp Thr Asp Phe Ser Lys Val Ala Ser Gly Lys Tyr Thr Ala Asn Tyr
130          135          140
Leu Asp Phe His Pro Asn Glu Leu His Ala Gly Asp Ser Gly Thr Phe

```

145					150					155				160
Gly	Gly	Tyr	Pro	Asp	Ile	Cys	His	Asp	Lys	Ser	Trp	Asp	Gln	Tyr
				165					170					175
Leu	Trp	Ala	Ser	Gln	Glu	Ser	Tyr	Ala	Ala	Tyr	Leu	Arg	Ser	Ile
			180					185					190	
Ile	Asp	Ala	Trp	Arg	Phe	Asp	Tyr	Val	Lys	Gly	Tyr	Gly	Ala	Trp
	195					200					205			
Val	Lys	Asp	Trp	Leu	Asn	Trp	Trp	Gly	Gly	Trp	Ala	Val	Gly	Glu
	210				215					220				
Trp	Asp	Thr	Asn	Val	Asp	Ala	Leu	Leu	Asn	Trp	Ala	Tyr	Ser	Ser
225					230					235				240
Ala	Lys	Val	Phe	Asp	Phe	Pro	Leu	Tyr	Tyr	Lys	Met	Asp	Ala	Ala
			245					250					255	
Asp	Asn	Lys	Asn	Ile	Pro	Ala	Leu	Val	Glu	Ala	Leu	Lys	Asn	Gly
		260						265					270	
Thr	Val	Val	Ser	Arg	Asp	Pro	Phe	Lys	Ala	Val	Thr	Phe	Val	Ala
	275						280					285		
His	Asp	Thr	Asp	Ile	Ile	Trp	Asn	Lys	Tyr	Pro	Ala	Tyr	Ala	Phe
	290				295					300				
Leu	Thr	Tyr	Glu	Gly	Gln	Pro	Thr	Ile	Phe	Tyr	Arg	Asp	Tyr	Glu
305					310					315				320
Trp	Leu	Asn	Lys	Asp	Lys	Leu	Lys	Asn	Leu	Ile	Trp	Ile	His	Asp
			325					330					335	
Leu	Ala	Gly	Gly	Ser	Thr	Ser	Ile	Val	Tyr	Tyr	Asp	Ser	Asp	Glu
	340						345					350		
Ile	Phe	Val	Arg	Asn	Gly	Tyr	Gly	Ser	Lys	Pro	Gly	Leu	Ile	Thr
	355				360					365				
Ile	Asn	Leu	Gly	Ser	Ser	Lys	Val	Gly	Arg	Trp	Val	Tyr	Val	Pro
	370				375					380				
Phe	Ala	Gly	Ala	Cys	Ile	His	Glu	Tyr	Thr	Gly	Asn	Leu	Gly	Gly
385				390					395					400
Val	Asp	Lys	Tyr	Val	Tyr	Ser	Ser	Gly	Trp	Val	Tyr	Leu	Glu	Ala
			405					410					415	
Ala	Tyr	Asp	Pro	Ala	Asn	Gly	Gln	Tyr	Gly	Tyr	Ser	Val	Trp	Ser
	420						425					430		
Cys	Gly	Val	Gly	Thr	Ser	Ile	Ala	Gly	Ile	Leu	Glu	Ala	Asp	Arg
	435					440					445			
Leu	Thr	Val	Ser	Pro	Tyr	Tyr	Ala	Glu	Glu	Leu	Ile	Ser	Gly	Ile
	450				455					460				
Arg	Gly	Cys	Glu	Leu	Asp	Asn	Ile	Met	Arg	Leu	Thr	Gly	Ile	Thr
465				470					475					480
Ile	Val	Asn	Gly	Met	Asp	Val	Ser	Glu	Trp	Asp	Pro	Ser	Arg	Asp
		485						490					495	
Tyr	Ile	Ala	Val	Lys	Tyr	Asp	Val	Ser	Thr	Ala	Val	Glu	Ala	Lys
	500					505						510		
Leu	Asn	Lys	Glu	Ala	Leu	Gln	Ala	Glu	Val	Gly	Leu	Pro	Val	Asp
	515					520					525			
Asn	Ile	Pro	Leu	Val	Ala	Phe	Ile	Gly	Arg	Leu	Glu	Glu	Gln	Lys
	530				535					540				
Pro	Asp	Val	Met	Ala	Ala	Ala	Ile	Pro	Gln	Leu	Met	Glu	Met	Val
545				550					555					560
Asp	Val	Gln	Ile	Val	Leu	Leu	Gly	Thr	Gly	Lys	Lys	Lys	Phe	Glu
			565					570					575	
Met	Leu	Met	Ser	Ala	Glu	Glu	Lys	Phe	Pro	Gly	Lys	Val	Arg	Ala

```

      580      585      590
Val Lys Phe Asn Ala Ala Leu Ala His His Ile Met Ala Gly Ala Asp
      595      600      605
Val Leu Ala Val Thr Ser Arg Phe Glu Pro Cys Gly Leu Ile Gln Leu
      610      615      620
Gln Gly Met Arg Tyr Gly Thr Pro Cys Ala Cys Ala Ser Thr Gly Gly
      625      630      635      640
Leu Val Asp Thr Ile Ile Glu Gly Lys Thr Gly Phe His Met Gly Arg
      645      650      655
Leu Ser Val Asp Cys Asn Val Val Glu Pro Ala Asp Val Lys Lys Val
      660      665      670
Ala Thr Thr Leu Gln Arg Ala Ile Lys Val Val Gly Thr Pro Ala Tyr
      675      680      685
Glu Glu Met Val Arg Asn Cys Met Ile Gln Asp Leu Ser Trp Lys Gly
      690      695      700
Pro Ala Lys Asn Trp Glu Asn Val Leu Leu Ser Leu Gly Val Ala Gly
      705      710      715      720
Gly Glu Pro Gly Val Glu Gly Glu Glu Ile Ala Pro Leu Ala Lys Glu
      725      730      735
Asn Val Ala Ala Pro
      740

```

<210> 11  
 <211> 1515  
 <212> DNA  
 <213> Zea mays

```

<400> 11
ggagagctat gagacgtatg tcctcaaagc cactttgcat tgtgtgaaac caatatcgat 60
ctttgttact tcatcatgca tgaacatttg tggaaactac tagcttacia gcattagtga 120
cagctcagaa aaaagttatc tatgaaaggt ttcattgtgta ccgtgggaaa tgagaaatgt 180
tgccaaactca aacaccttca atatgttgtt tgcaggcaaa ctcttctgga agaaaggtgt 240
ctaaaaactat gaacgggtta cagaaaggta taaaccacgg ctgtgcattt tggaaagtac 300
atctatagat gtctgttgag gggaaagccg tacgccaacg ttatttactc agaaacagct 360
tcaacacaca gttgtctgct ttatgatggc atctccaccc aggcacccac catcacctat 420
ctctcgtgcc tgtttatttt cttgcccttt ctgatcataa aaaaacatta agagtttgca 480
aacatgcata ggcatatcaa tatgtctcatt tattaatttg ctacgagatc atcttcctac 540
tctttacttt atttattgtt tgaaaaatat gtccctgcacc tagggagctc gtatacagta 600
ccaatgcata ttcattaaat gtgaatttca gaaaggaaat aggaacctat gagagtattt 660
ttcaaaatta attagcggct tctattatgt ttatagcaaa ggccaagggc aaaattggaa 720
cactaatgat ggttggttgc atgagtcctgt cgattacttg caagaaatgt gaacctttgt 780
ttctgtgcgt gggcataaaa caaacagctt ctacgctctt ttacgggtact tgcacttgca 840
agaaatgtga actccttttc atttctgtat gtggacataa tgccaaagca tccaggcttt 900
ttcatggttg ttgatgtctt tacacagttc atctccacca gtatgccctc ctcatactct 960
atataaacac atcaacagca tcgcaattag ccacaagatc acttcgggag gcaagtgcga 1020
tttcgatctc gcagccacct tttttgttgc tgttgtaagt ataccttccc ttaccatctt 1080
tatctgttag tttaatttgt aattgggaag tattagtgga aagaggatga gatgctatca 1140
tctatgtact ctgcaaatgc atctgacgtt atatgggctg cttcatataa tttgaattgc 1200
tccattcttg ccgacaatat attgcaaggt atatgcctag ttccatcaaa agttctgttt 1260
tttcattcta aaagcatttt agtggcacac aatttttgtc catgagggaa aggaaatctg 1320
ttttggttac tttgtttgag gtgcattctt catatgtcca gttttatgga agtaataaac 1380
ttcagtttgg tcataagatg tcatataaaa gggcaaacat atattcaatg ttcaattcat 1440
cgtaaatggt ccctttttgt aaaagattgc atactcattt atttgagttg cagggtgtatc 1500

```



tagtagttgg aggag

1515

&lt;210&gt; 12

&lt;211&gt; 673

&lt;212&gt; DNA

&lt;213&gt; Zea mays

&lt;400&gt; 12

```

gatcatccag gtgcaaccgt ataagtccta aagtgggtgag gaacacgaaa caaccatgca 60
ttggcatgta aagctccaag aatttggtgt atccttaaca actcacagaa catcaaccaa 120
aattgcacgt caaggggtatt gggttaagaaa caatcaaaca aatcctctct gtgtgcaaag 180
aaacacgggtg agtcatgccg agatcatact catctgatat acatgcttac agctcacaag 240
acattacaaa caactcatat tgcattacaa agatcggttc atgaaaaata aaataggccg 300
gacaggacaa aaatccttga cgtgtaaagt aaatttataa caaaaaaaaa gccatatgtc 360
aagctaaatc taattcgttt tacgtagatc aacaacctgt agaaggcaac aaaactgagc 420
cacgcagaag tacagaatga ttccagatga accatcgacg tgctacgtaa agagagtgac 480
gagtcataata catttggcaa gaaacatga agctgcctac agccgtctcg gtggcataag 540
aacacaagaa attgtgttaa ttaatcaaag ctataaataa cgctcgcatg cctgtgcact 600
tctccatcac caccactggg tcttcagacc attagcttta tctactccag agcgcagaag 660
aaccgatcg aca 673

```

&lt;210&gt; 13

&lt;211&gt; 454

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic

&lt;400&gt; 13

```

Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser
 1           5           10          15
Ala Thr Ser Ala Lys Tyr Leu Glu Leu Glu Gly Gly Val Ile Met
 20          25          30
Gln Ala Phe Tyr Trp Asp Val Pro Ser Gly Gly Ile Trp Trp Asp Thr
 35          40          45
Ile Arg Gln Lys Ile Pro Glu Trp Tyr Asp Ala Gly Ile Ser Ala Ile
 50          55          60
Trp Ile Pro Pro Ala Ser Lys Gly Met Ser Gly Gly Tyr Ser Met Gly
 65          70          75          80
Tyr Asp Pro Tyr Asp Tyr Phe Asp Leu Gly Glu Tyr Tyr Gln Lys Gly
 85          90          95
Thr Val Glu Thr Arg Phe Gly Ser Lys Gln Glu Leu Ile Asn Met Ile
100         105         110
Asn Thr Ala His Ala Tyr Gly Ile Lys Val Ile Ala Asp Ile Val Ile
115         120         125
Asn His Arg Ala Gly Gly Asp Leu Glu Trp Asn Pro Phe Val Gly Asp
130         135         140
Tyr Thr Trp Thr Asp Phe Ser Lys Val Ala Ser Gly Lys Tyr Thr Ala
145         150         155         160
Asn Tyr Leu Asp Phe His Pro Asn Glu Leu His Ala Gly Asp Ser Gly
165         170         175
Thr Phe Gly Gly Tyr Pro Asp Ile Cys His Asp Lys Ser Trp Asp Gln
180         185         190

```

Tyr Trp Leu Trp Ala Ser Gln Glu Ser Tyr Ala Ala Tyr Leu Arg Ser  
                   195                  200                  205  
 Ile Gly Ile Asp Ala Trp Arg Phe Asp Tyr Val Lys Gly Tyr Gly Ala  
                   210                  215                  220  
 Trp Val Val Lys Asp Trp Leu Asn Trp Trp Gly Gly Trp Ala Val Gly  
 225                  230                  235                  240  
 Glu Tyr Trp Asp Thr Asn Val Asp Ala Leu Leu Asn Trp Ala Tyr Ser  
                   245                  250                  255  
 Ser Gly Ala Lys Val Phe Asp Phe Pro Leu Tyr Tyr Lys Met Asp Ala  
                   260                  265                  270  
 Ala Phe Asp Asn Lys Asn Ile Pro Ala Leu Val Glu Ala Leu Lys Asn  
                   275                  280                  285  
 Gly Gly Thr Val Val Ser Arg Asp Pro Phe Lys Ala Val Thr Phe Val  
                   290                  295                  300  
 Ala Asn His Asp Thr Asp Ile Ile Trp Asn Lys Tyr Pro Ala Tyr Ala  
 305                  310                  315                  320  
 Phe Ile Leu Thr Tyr Glu Gly Gln Pro Thr Ile Phe Tyr Arg Asp Tyr  
                   325                  330                  335  
 Glu Glu Trp Leu Asn Lys Asp Lys Leu Lys Asn Leu Ile Trp Ile His  
                   340                  345                  350  
 Asp Asn Leu Ala Gly Gly Ser Thr Ser Ile Val Tyr Tyr Asp Ser Asp  
                   355                  360                  365  
 Glu Met Ile Phe Val Arg Asn Gly Tyr Gly Ser Lys Pro Gly Leu Ile  
                   370                  375                  380  
 Thr Tyr Ile Asn Leu Gly Ser Ser Lys Val Gly Arg Trp Val Tyr Val  
 385                  390                  395                  400  
 Pro Lys Phe Ala Gly Ala Cys Ile His Glu Tyr Thr Gly Asn Leu Gly  
                   405                  410                  415  
 Gly Trp Val Asp Lys Tyr Val Tyr Ser Ser Gly Trp Val Tyr Leu Glu  
                   420                  425                  430  
 Ala Pro Ala Tyr Asp Pro Ala Asn Gly Gln Tyr Gly Tyr Ser Val Trp  
                   435                  440                  445  
 Ser Tyr Cys Gly Val Gly  
                   450

&lt;210&gt; 14

&lt;211&gt; 460

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic

&lt;400&gt; 14

Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser  
                   5                  10                  15  
 Ala Thr Ser Ala Lys Tyr Leu Glu Leu Glu Glu Gly Gly Val Ile Met  
                   20                  25                  30  
 Gln Ala Phe Tyr Trp Asp Val Pro Ser Gly Gly Ile Trp Trp Asp Thr  
                   35                  40                  45  
 Ile Arg Gln Lys Ile Pro Glu Trp Tyr Asp Ala Gly Ile Ser Ala Ile  
                   50                  55                  60  
 Trp Ile Pro Pro Ala Ser Lys Gly Met Ser Gly Gly Tyr Ser Met Gly

```

65          70          75          80
Tyr Asp Pro Tyr Asp Tyr Phe Asp Leu Gly Glu Tyr Tyr Gln Lys Gly
      85          90          95
Thr Val Glu Thr Arg Phe Gly Ser Lys Gln Glu Leu Ile Asn Met Ile
      100          105          110
Asn Thr Ala His Ala Tyr Gly Ile Lys Val Ile Ala Asp Ile Val Ile
      115          120          125
Asn His Arg Ala Gly Gly Asp Leu Glu Trp Asn Pro Phe Val Gly Asp
      130          135          140
Tyr Thr Trp Thr Asp Phe Ser Lys Val Ala Ser Gly Lys Tyr Thr Ala
145          150          155          160
Asn Tyr Leu Asp Phe His Pro Asn Glu Leu His Ala Gly Asp Ser Gly
      165          170          175
Thr Phe Gly Gly Tyr Pro Asp Ile Cys His Asp Lys Ser Trp Asp Gln
      180          185          190
Tyr Trp Leu Trp Ala Ser Gln Glu Ser Tyr Ala Ala Tyr Leu Arg Ser
      195          200          205
Ile Gly Ile Asp Ala Trp Arg Phe Asp Tyr Val Lys Gly Tyr Gly Ala
210          215          220
Trp Val Val Lys Asp Trp Leu Asn Trp Trp Gly Gly Trp Ala Val Gly
225          230          235          240
Glu Tyr Trp Asp Thr Asn Val Asp Ala Leu Leu Asn Trp Ala Tyr Ser
      245          250          255
Ser Gly Ala Lys Val Phe Asp Phe Pro Leu Tyr Tyr Lys Met Asp Ala
      260          265          270
Ala Phe Asp Asn Lys Asn Ile Pro Ala Leu Val Glu Ala Leu Lys Asn
      275          280          285
Gly Gly Thr Val Val Ser Arg Asp Pro Phe Lys Ala Val Thr Phe Val
290          295          300
Ala Asn His Asp Thr Asp Ile Ile Trp Asn Lys Tyr Pro Ala Tyr Ala
305          310          315          320
Phe Ile Leu Thr Tyr Glu Gly Gln Pro Thr Ile Phe Tyr Arg Asp Tyr
      325          330          335
Glu Glu Trp Leu Asn Lys Asp Lys Leu Lys Asn Leu Ile Trp Ile His
      340          345          350
Asp Asn Leu Ala Gly Gly Ser Thr Ser Ile Val Tyr Tyr Asp Ser Asp
      355          360          365
Glu Met Ile Phe Val Arg Asn Gly Tyr Gly Ser Lys Pro Gly Leu Ile
      370          375          380
Thr Tyr Ile Asn Leu Gly Ser Ser Lys Val Gly Arg Trp Val Tyr Val
385          390          395          400
Pro Lys Phe Ala Gly Ala Cys Ile His Glu Tyr Thr Gly Asn Leu Gly
      405          410          415
Gly Trp Val Asp Lys Tyr Val Tyr Ser Ser Gly Trp Val Tyr Leu Glu
      420          425          430
Ala Pro Ala Tyr Asp Pro Ala Asn Gly Gln Tyr Gly Tyr Ser Val Trp
      435          440          445
Ser Tyr Cys Gly Val Gly Ser Glu Lys Asp Glu Leu
450          455          460

```

<210> 15  
<211> 518  
<212> PRT

## &lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic

&lt;400&gt; 15

```

Met Leu Ala Ala Leu Ala Thr Ser Gln Leu Val Ala Thr Arg Ala Gly
 1           5           10           15
Leu Gly Val Pro Asp Ala Ser Thr Phe Arg Arg Gly Ala Ala Gln Gly
      20           25           30
Leu Arg Gly Ala Arg Ala Ser Ala Ala Ala Asp Thr Leu Ser Met Arg
      35           40           45
Thr Ser Ala Arg Ala Ala Pro Arg His Gln His Gln Gln Ala Arg Arg
      50           55           60
Gly Ala Arg Phe Pro Ser Leu Val Val Cys Ala Ser Ala Gly Ala Met
 65           70           75           80
Ala Lys Tyr Leu Glu Leu Glu Gly Gly Val Ile Met Gln Ala Phe
      85           90           95
Tyr Trp Asp Val Pro Ser Gly Gly Ile Trp Trp Asp Thr Ile Arg Gln
      100          105          110
Lys Ile Pro Glu Trp Tyr Asp Ala Gly Ile Ser Ala Ile Trp Ile Pro
      115          120          125
Pro Ala Ser Lys Gly Met Ser Gly Gly Tyr Ser Met Gly Tyr Asp Pro
      130          135          140
Tyr Asp Tyr Phe Asp Leu Gly Glu Tyr Tyr Gln Lys Gly Thr Val Glu
 145          150          155          160
Thr Arg Phe Gly Ser Lys Gln Glu Leu Ile Asn Met Ile Asn Thr Ala
      165          170          175
His Ala Tyr Gly Ile Lys Val Ile Ala Asp Ile Val Ile Asn His Arg
      180          185          190
Ala Gly Gly Asp Leu Glu Trp Asn Pro Phe Val Gly Asp Tyr Thr Trp
      195          200          205
Thr Asp Phe Ser Lys Val Ala Ser Gly Lys Tyr Thr Ala Asn Tyr Leu
      210          215          220
Asp Phe His Pro Asn Glu Leu His Ala Gly Asp Ser Gly Thr Phe Gly
 225          230          235          240
Gly Tyr Pro Asp Ile Cys His Asp Lys Ser Trp Asp Gln Tyr Trp Leu
      245          250          255
Trp Ala Ser Gln Glu Ser Tyr Ala Ala Tyr Leu Arg Ser Ile Gly Ile
      260          265          270
Asp Ala Trp Arg Phe Asp Tyr Val Lys Gly Tyr Gly Ala Trp Val Val
      275          280          285
Lys Asp Trp Leu Asn Trp Trp Gly Gly Trp Ala Val Gly Glu Tyr Trp
      290          295          300
Asp Thr Asn Val Asp Ala Leu Leu Asn Trp Ala Tyr Ser Ser Gly Ala
 305          310          315          320
Lys Val Phe Asp Phe Pro Leu Tyr Tyr Lys Met Asp Ala Ala Phe Asp
      325          330          335
Asn Lys Asn Ile Pro Ala Leu Val Glu Ala Leu Lys Asn Gly Gly Thr
      340          345          350
Val Val Ser Arg Asp Pro Phe Lys Ala Val Thr Phe Val Ala Asn His
      355          360          365
Asp Thr Asp Ile Ile Trp Asn Lys Tyr Pro Ala Tyr Ala Phe Ile Leu
 370          375          380

```

Thr Tyr Glu Gly Gln Pro Thr Ile Phe Tyr Arg Asp Tyr Glu Glu Trp  
 385 390 395 400  
 Leu Asn Lys Asp Lys Leu Lys Asn Leu Ile Trp Ile His Asp Asn Leu  
 405 410 415  
 Ala Gly Gly Ser Thr Ser Ile Val Tyr Tyr Asp Ser Asp Glu Met Ile  
 420 425 430  
 Phe Val Arg Asn Gly Tyr Gly Ser Lys Pro Gly Leu Ile Thr Tyr Ile  
 435 440 445  
 Asn Leu Gly Ser Ser Lys Val Gly Arg Trp Val Tyr Val Pro Lys Phe  
 450 455 460  
 Ala Gly Ala Cys Ile His Glu Tyr Thr Gly Asn Leu Gly Gly Trp Val  
 465 470 475 480  
 Asp Lys Tyr Val Tyr Ser Ser Gly Trp Val Tyr Leu Glu Ala Pro Ala  
 485 490 495  
 Tyr Asp Pro Ala Asn Gly Gln Tyr Gly Tyr Ser Val Trp Ser Tyr Cys  
 500 505 510  
 Gly Val Gly Thr Ser Ile  
 515

&lt;210&gt; 16

&lt;211&gt; 820

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic

&lt;400&gt; 16

Met Leu Ala Ala Leu Ala Thr Ser Gln Leu Val Ala Thr Arg Ala Gly  
 1 5 10 15  
 Leu Gly Val Pro Asp Ala Ser Thr Phe Arg Arg Gly Ala Ala Gln Gly  
 20 25 30  
 Leu Arg Gly Ala Arg Ala Ser Ala Ala Asp Thr Leu Ser Met Arg  
 35 40 45  
 Thr Ser Ala Arg Ala Ala Pro Arg His Gln His Gln Gln Ala Arg Arg  
 50 55 60  
 Gly Ala Arg Phe Pro Ser Leu Val Val Cys Ala Ser Ala Gly Ala Met  
 65 70 75 80  
 Ala Lys Tyr Leu Glu Leu Glu Gly Gly Val Ile Met Gln Ala Phe  
 85 90 95  
 Tyr Trp Asp Val Pro Ser Gly Gly Ile Trp Trp Asp Thr Ile Arg Gln  
 100 105 110  
 Lys Ile Pro Glu Trp Tyr Asp Ala Gly Ile Ser Ala Ile Trp Ile Pro  
 115 120 125  
 Pro Ala Ser Lys Gly Met Ser Gly Gly Tyr Ser Met Gly Tyr Asp Pro  
 130 135 140  
 Tyr Asp Tyr Phe Asp Leu Gly Glu Tyr Tyr Gln Lys Gly Thr Val Glu  
 145 150 155 160  
 Thr Arg Phe Gly Ser Lys Gln Glu Leu Ile Asn Met Ile Asn Thr Ala  
 165 170 175  
 His Ala Tyr Gly Ile Lys Val Ile Ala Asp Ile Val Ile Asn His Arg  
 180 185 190  
 Ala Gly Gly Asp Leu Glu Trp Asn Pro Phe Val Gly Asp Tyr Thr Trp

195	200	205
Thr Asp Phe Ser Lys Val	Ala Ser Gly Lys Tyr	Thr Ala Asn Tyr Leu
210	215	220
Asp Phe His Pro Asn Glu	Leu His Ala Gly Asp	Ser Gly Thr Phe Gly
225	230	235
Gly Tyr Pro Asp Ile Cys	His Asp Lys Ser Trp	Asp Gln Tyr Trp Leu
245	250	255
Trp Ala Ser Gln Glu Ser	Tyr Ala Ala Tyr Leu	Arg Ser Ile Gly Ile
260	265	270
Asp Ala Trp Arg Phe Asp	Tyr Val Lys Gly Tyr Gly	Ala Trp Val Val
275	280	285
Lys Asp Trp Leu Asn Trp	Trp Gly Gly Trp Ala Val	Gly Glu Tyr Trp
290	295	300
Asp Thr Asn Val Asp Ala	Leu Leu Asn Trp Ala Tyr	Ser Ser Gly Ala
305	310	315
Lys Val Phe Asp Phe Pro	Leu Tyr Tyr Lys Met	Asp Ala Ala Phe Asp
325	330	335
Asn Lys Asn Ile Pro Ala	Leu Val Glu Ala Leu Lys	Asn Gly Gly Thr
340	345	350
Val Val Ser Arg Asp Pro	Phe Lys Ala Val Thr Phe	Val Ala Asn His
355	360	365
Asp Thr Asp Ile Ile Trp	Asn Lys Tyr Pro Ala Tyr	Ala Phe Ile Leu
370	375	380
Thr Tyr Glu Gly Gln Pro	Thr Ile Phe Tyr Arg Asp	Tyr Glu Glu Trp
385	390	395
Leu Asn Lys Asp Lys Leu	Lys Asn Leu Ile Trp Ile	His Asp Asn Leu
405	410	415
Ala Gly Gly Ser Thr Ser	Ile Val Tyr Tyr Asp Ser	Asp Glu Met Ile
420	425	430
Phe Val Arg Asn Gly Tyr	Gly Ser Lys Pro Gly Leu	Ile Thr Tyr Ile
435	440	445
Asn Leu Gly Ser Ser Lys	Val Gly Arg Trp Val Tyr	Val Pro Lys Phe
450	455	460
Ala Gly Ala Cys Ile His	Glu Tyr Thr Gly Asn Leu	Gly Gly Trp Val
465	470	475
Asp Lys Tyr Val Tyr Ser	Ser Gly Trp Val Tyr Leu	Glu Ala Pro Ala
485	490	495
Tyr Asp Pro Ala Asn Gly	Gln Tyr Gly Tyr Ser Val	Trp Ser Tyr Cys
500	505	510
Gly Val Gly Thr Ser Ile	Ala Gly Ile Leu Glu Ala	Asp Arg Val Leu
515	520	525
Thr Val Ser Pro Tyr Tyr	Ala Glu Glu Leu Ile Ser	Gly Ile Ala Arg
530	535	540
Gly Cys Glu Leu Asp Asn	Ile Met Arg Leu Thr Gly	Ile Thr Gly Ile
545	550	555
Val Asn Gly Met Asp Val	Ser Glu Trp Asp Pro Ser	Arg Asp Lys Tyr
565	570	575
Ile Ala Val Lys Tyr Asp	Val Ser Thr Ala Val Glu	Ala Lys Ala Leu
580	585	590
Asn Lys Glu Ala Leu Gln	Ala Glu Val Gly Leu Pro	Val Asp Arg Asn
595	600	605
Ile Pro Leu Val Ala Phe	Ile Gly Arg Leu Glu Glu	Gln Lys Gly Pro
610	615	620
Asp Val Met Ala Ala Ala	Ile Pro Gln Leu Met Glu	Met Val Glu Asp

```

625          630          635          640
Val Gln Ile Val Leu Leu Gly Thr Gly Lys Lys Lys Phe Glu Arg Met
          645          650          655
Leu Met Ser Ala Glu Glu Lys Phe Pro Gly Lys Val Arg Ala Val Val
          660          665          670
Lys Phe Asn Ala Ala Leu Ala His His Ile Met Ala Gly Ala Asp Val
          675          680          685
Leu Ala Val Thr Ser Arg Phe Glu Pro Cys Gly Leu Ile Gln Leu Gln
          690          695          700
Gly Met Arg Tyr Gly Thr Pro Cys Ala Cys Ala Ser Thr Gly Gly Leu
705          710          715          720
Val Asp Thr Ile Ile Glu Gly Lys Thr Gly Phe His Met Gly Arg Leu
          725          730          735
Ser Val Asp Cys Asn Val Val Glu Pro Ala Asp Val Lys Lys Val Ala
          740          745          750
Thr Thr Leu Gln Arg Ala Ile Lys Val Val Gly Thr Pro Ala Tyr Glu
          755          760          765
Glu Met Val Arg Asn Cys Met Ile Gln Asp Leu Ser Trp Lys Gly Pro
          770          775          780
Ala Lys Asn Trp Glu Asn Val Leu Leu Ser Leu Gly Val Ala Gly Gly
785          790          795          800
Glu Pro Gly Val Glu Gly Glu Glu Ile Ala Pro Leu Ala Lys Glu Asn
          805          810          815
Val Ala Ala Pro
          820

```

```

<210> 17
<211> 19
<212> PRT
<213> Artificial Sequence

```

```

<220>
<223> synthetic

```

```

<400> 17
Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser
1          5          10          15
Ala Thr Ser

```

```

<210> 18
<211> 444
<212> PRT
<213> Thermotoga maritima

```

```

<400> 18
Met Ala Glu Phe Phe Pro Glu Ile Pro Lys Ile Gln Phe Glu Gly Lys
1          5          10          15
Glu Ser Thr Asn Pro Leu Ala Phe Arg Phe Tyr Asp Pro Asn Glu Val
          20          25          30
Ile Asp Gly Lys Pro Leu Lys Asp His Leu Lys Phe Ser Val Ala Phe
          35          40          45

```

```

Trp His Thr Phe Val Asn Glu Gly Arg Asp Pro Phe Gly Asp Pro Thr
  50                      55                      60
Ala Glu Arg Pro Trp Asn Arg Phe Ser Asp Pro Met Asp Lys Ala Phe
  65                      70                      75                      80
Ala Arg Val Asp Ala Leu Phe Glu Phe Cys Glu Lys Leu Asn Ile Glu
                      85                      90                      95
Tyr Phe Cys Phe His Asp Arg Asp Ile Ala Pro Glu Gly Lys Thr Leu
                      100                      105                      110
Arg Glu Thr Asn Lys Ile Leu Asp Lys Val Val Glu Arg Ile Lys Glu
                      115                      120                      125
Arg Met Lys Asp Ser Asn Val Lys Leu Leu Trp Gly Thr Ala Asn Leu
                      130                      135                      140
Phe Ser His Pro Arg Tyr Met His Gly Ala Ala Thr Thr Cys Ser Ala
  145                      150                      155                      160
Asp Val Phe Ala Tyr Ala Ala Ala Gln Val Lys Lys Ala Leu Glu Ile
                      165                      170                      175
Thr Lys Glu Leu Gly Gly Glu Gly Tyr Val Phe Trp Gly Gly Arg Glu
                      180                      185                      190
Gly Tyr Glu Thr Leu Leu Asn Thr Asp Leu Gly Leu Glu Leu Glu Asn
                      195                      200                      205
Leu Ala Arg Phe Leu Arg Met Ala Val Glu Tyr Ala Lys Lys Ile Gly
  210                      215                      220
Phe Thr Gly Gln Phe Leu Ile Glu Pro Lys Pro Lys Glu Pro Thr Lys
  225                      230                      235                      240
His Gln Tyr Asp Phe Asp Val Ala Thr Ala Tyr Ala Phe Leu Lys Asn
                      245                      250                      255
His Gly Leu Asp Glu Tyr Phe Lys Phe Asn Ile Glu Ala Asn His Ala
                      260                      265                      270
Thr Leu Ala Gly His Thr Phe Gln His Glu Leu Arg Met Ala Arg Ile
                      275                      280                      285
Leu Gly Lys Leu Gly Ser Ile Asp Ala Asn Gln Gly Asp Leu Leu Leu
  290                      295                      300
Gly Trp Asp Thr Asp Gln Phe Pro Thr Asn Ile Tyr Asp Thr Thr Leu
  305                      310                      315                      320
Ala Met Tyr Glu Val Ile Lys Ala Gly Gly Phe Thr Lys Gly Gly Leu
                      325                      330                      335
Asn Phe Asp Ala Lys Val Arg Arg Ala Ser Tyr Lys Val Glu Asp Leu
                      340                      345                      350
Phe Ile Gly His Ile Ala Gly Met Asp Thr Phe Ala Leu Gly Phe Lys
                      355                      360                      365
Ile Ala Tyr Lys Leu Ala Lys Asp Gly Val Phe Asp Lys Phe Ile Glu
  370                      375                      380
Glu Lys Tyr Arg Ser Phe Lys Glu Gly Ile Gly Lys Glu Ile Val Glu
  385                      390                      395                      400
Gly Lys Thr Asp Phe Glu Lys Leu Glu Glu Tyr Ile Ile Asp Lys Glu
                      405                      410                      415
Asp Ile Glu Leu Pro Ser Gly Lys Gln Glu Tyr Leu Glu Ser Leu Leu
                      420                      425                      430
Asn Ser Tyr Ile Val Lys Thr Ile Ala Glu Leu Arg
                      435                      440

```

&lt;210&gt; 19

&lt;211&gt; 1335



&lt;212&gt; DNA

<213> *Thermotoga maritima*

&lt;400&gt; 19

```

atggccgagt tcttcccgga gatcccgaag atccagttcg agggcaagga gtccaccaac 60
ccgctcgctt tccgcttcta cgacccgaac gaggtgatcg acggcaagcc gctcaaggac 120
cacctcaagt tctccgtggc cttctggcac accttcgtga acgagggccg cgacccgttc 180
ggcgacccga ccgccgagcg cccgtggaac cgcttctccg acccgatgga caaggccttc 240
gcccgcgtgg acgccctctt cgagttctgc gagaagctca acatcgagta cttctgcttc 300
cacgaccgcg acatcgcccc ggagggcaag accctccgcg agaccaacaa gatcctcgac 360
aaggtggtgg agcgcataaa ggagcgcatg aaggactcca acgtgaagct cctctggggc 420
accgccaacc tcttctccca ccgcgctac atgcacggcg ccgccaccac ctgctccgcc 480
gacgtgttcg cctacgccgc cgcccaggtg aagaaggccc tggagatcac caaggagctg 540
ggcgccgagg gctacgtgtt ctggggcggc cgcgagggct acgagaccct cctcaacacc 600
gacctcggcc tggagctgga gaacctcgcc cgcttctccc gcatggccgt ggagtacgcc 660
aagaagatcg gcttcaccgg ccagttcttc atcgagccga agccgaagga gccgaccaag 720
caccagtacg acttcgacgt ggccaccgcc tacgccttcc tcaagaacca cggcctcgac 780
gagtacttca agttcaacat cgaggccaac caccgccacc tcgccggcca cacttccag 840
cacgagctgc gcatggcccc catectcggc aagctcggct ccatcgacgc caaccagggc 900
gacctctccc tcggctggga caccgaccag ttcccaccca acatctacga caccaccctc 960
gccatgtacg aggtgatcaa ggccggcggc ttaccaagg gcggcctcaa cttcgacgcc 1020
aaggtgcgcc gcgcctccta caaggtggag gacctcttca tcggccacat cgccggcatg 1080
gacaccttcg cctcgggctt caagatcgcc tacaagctcg ccaaggacgg cgtgttcgac 1140
aagttcatcg aggagaagta ccgctccttc aaggagggca tcggcaagga gatcgtggag 1200
ggcaagaccg acttcgagaa gctggaggag tacatcatcg acaaggagga catcgagctg 1260
ccgtccggca agcaggagta cctggagtcc ctcctcaact cctacatcgt gaagaccatc 1320
gccgagctgc gctga                                     1335

```

&lt;210&gt; 20

&lt;211&gt; 444

&lt;212&gt; PRT

<213> *Thermotoga neapolitana*

&lt;400&gt; 20

```

Met Ala Glu Phe Phe Pro Glu Ile Pro Lys Val Gln Phe Glu Gly Lys
 1           5           10          15
Glu Ser Thr Asn Pro Leu Ala Phe Lys Phe Tyr Asp Pro Glu Glu Ile
          20          25          30
Ile Asp Gly Lys Pro Leu Lys Asp His Leu Lys Phe Ser Val Ala Phe
          35          40          45
Trp His Thr Phe Val Asn Glu Gly Arg Asp Pro Phe Gly Asp Pro Thr
          50          55          60
Ala Asp Arg Pro Trp Asn Arg Tyr Thr Asp Pro Met Asp Lys Ala Phe
          65          70          75          80
Ala Arg Val Asp Ala Leu Phe Glu Phe Cys Glu Lys Leu Asn Ile Glu
          85          90          95
Tyr Phe Cys Phe His Asp Arg Asp Ile Ala Pro Glu Gly Lys Thr Leu
          100         105         110
Arg Glu Thr Asn Lys Ile Leu Asp Lys Val Val Glu Arg Ile Lys Glu
          115         120         125
Arg Met Lys Asp Ser Asn Val Lys Leu Leu Trp Gly Thr Ala Asn Leu
          130         135         140
Phe Ser His Pro Arg Tyr Met His Gly Ala Ala Thr Thr Cys Ser Ala
          145         150         155         160

```

Asp Val Phe Ala Tyr Ala Ala Ala Gln Val Lys Lys Ala Leu Glu Ile  
 165 170 175  
 Thr Lys Glu Leu Gly Gly Glu Gly Tyr Val Phe Trp Gly Gly Arg Glu  
 180 185 190  
 Gly Tyr Glu Thr Leu Leu Asn Thr Asp Leu Gly Phe Glu Leu Glu Asn  
 195 200 205  
 Leu Ala Arg Phe Leu Arg Met Ala Val Asp Tyr Ala Lys Arg Ile Gly  
 210 215 220  
 Phe Thr Gly Gln Phe Leu Ile Glu Pro Lys Pro Lys Glu Pro Thr Lys  
 225 230 235 240  
 His Gln Tyr Asp Phe Asp Val Ala Thr Ala Tyr Ala Phe Leu Lys Ser  
 245 250 255  
 His Gly Leu Asp Glu Tyr Phe Lys Phe Asn Ile Glu Ala Asn His Ala  
 260 265 270  
 Thr Leu Ala Gly His Thr Phe Gln His Glu Leu Arg Met Ala Arg Ile  
 275 280 285  
 Leu Gly Lys Leu Gly Ser Ile Asp Ala Asn Gln Gly Asp Leu Leu Leu  
 290 295 300  
 Gly Trp Asp Thr Asp Gln Phe Pro Thr Asn Val Tyr Asp Thr Thr Leu  
 305 310 315 320  
 Ala Met Tyr Glu Val Ile Lys Ala Gly Gly Phe Thr Lys Gly Gly Leu  
 325 330 335  
 Asn Phe Asp Ala Lys Val Arg Arg Ala Ser Tyr Lys Val Glu Asp Leu  
 340 345 350  
 Phe Ile Gly His Ile Ala Gly Met Asp Thr Phe Ala Leu Gly Phe Lys  
 355 360 365  
 Val Ala Tyr Lys Leu Val Lys Asp Gly Val Leu Asp Lys Phe Ile Glu  
 370 375 380  
 Glu Lys Tyr Arg Ser Phe Arg Glu Gly Ile Gly Arg Asp Ile Val Glu  
 385 390 395 400  
 Gly Lys Val Asp Phe Glu Lys Leu Glu Glu Tyr Ile Ile Asp Lys Glu  
 405 410 415  
 Thr Ile Glu Leu Pro Ser Gly Lys Gln Glu Tyr Leu Glu Ser Leu Ile  
 420 425 430  
 Asn Ser Tyr Ile Val Lys Thr Ile Leu Glu Leu Arg  
 435 440

&lt;210&gt; 21

&lt;211&gt; 1335

&lt;212&gt; DNA

<213> *Thermotoga neapolitana*

&lt;400&gt; 21

atggccgagt tcttcccga gatcccgaag gtgcagttcg agggcaagga gtccaccaac 60  
 ccgtcgcct tcaagttcta cgaccggag gagatcatcg acggcaagcc gctcaaggac 120  
 cacctcaagt tctccgtggc cttctggcac accttcgtga acgagggccg cgaccggttc 180  
 ggcgacccga ccgccgaccg cccgtggaac cgctacaccg acccgatgga caaggccttc 240  
 gccgcgtgg acgcctctt cgagttctgc gagaagctca acatcgagta cttctgcttc 300  
 caccgaccg acatcgcccc ggagggcaag accctccgcg agaccaacaa gacccctgac 360  
 aagggtggtg agcgcacaa ggagcgcacg aaggactcca acgtgaagct cctctggggc 420  
 accgccaacc tcttctccca ccgcgctac atgcacggcg ccgccaccac ctgctccgac 480  
 gacgtgttcg cctacgcgcg cgcccaggtg aagaaggccc tggagatcac caaggagctg 540  
 ggccgagagg gctacgtgtt ctggggcgcc cgcgagggct acgagaccct cctcaacacc 600

```

gacctcggct tgcagctgga gaacctcgcc cgcttcctcc gcatggccgt ggactacgcc 660
aagcgcacgc gcttcaccgg ccagttcctc atcgagccga agccgaagga gccgaccaag 720
caccagtaac acttcgacgt ggccaccgcc tacgccttcc tcaagtccca cggcctcgac 780
gagtacttca agttcaacat cgaggccaac cagccacccc tcgccggcca caccttccag 840
cacgagctgc gcatggcccc catcctcggc aagctcggct ccatcgacgc caaccagggc 900
gacctctccc tcggctggga caccgaccag ttcccgaacca acgtgtacga caccaccctc 960
gccatgtacg aggtgatcaa ggccggcggc ttcaccaagg gcggcctcaa ctccgacgcc 1020
aaggtgcgcc gcgcctccta caaggtggag gacctcttca tcggccacat cgcgggcgat 1080
gacaccttcg cctcggcgtt caaggtggcc tacaagctcg tgaaggacgg cgtgctcgac 1140
aagttcatcg aggagaagta ccgctccttc cgcgagggca tcggccgcga catcgtggag 1200
ggcaagggtg acttcgagaa gctggaggag tacatcatcg acaaggagac catcgagctg 1260
ccgtccggca agcaggagta cctggagtcc ctcatcaact cctacatcgt gaagaccatc 1320
ctggagctgc gctga                                     1335

```

<210> 22  
 <211> 28  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 22  
 agcgaattca tggcggctct ggccacgt 28

<210> 23  
 <211> 29  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 23  
 agctaagctt cagggcgcgg ccacgttct 29

<210> 24  
 <211> 825  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 24  
 Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser  
 1 5 10 15  
 Ala Thr Ser Ala Gly His Trp Tyr Lys His Gln Arg Ala Tyr Gln Phe  
 20 25 30  
 Thr Gly Glu Asp Asp Phe Gly Lys Val Ala Val Val Lys Leu Pro Met  
 35 40 45  
 Asp Leu Thr Lys Val Gly Ile Ile Val Arg Leu Asn Glu Trp Gln Ala  
 50 55 60  
 Lys Asp Val Ala Lys Asp Arg Phe Ile Glu Ile Lys Asp Gly Lys Ala

65					70					75				80	
Glu	Val	Trp	Ile	Leu	Gln	Gly	Val	Glu	Glu	Ile	Phe	Tyr	Glu	Lys	Pro
				85					90					95	
Asp	Thr	Ser	Pro	Arg	Ile	Phe	Phe	Ala	Gln	Ala	Arg	Ser	Asn	Lys	Val
			100					105						110	
Ile	Glu	Ala	Phe	Leu	Thr	Asn	Pro	Val	Asp	Thr	Lys	Lys	Lys	Glu	Leu
		115					120					125			
Phe	Lys	Val	Thr	Val	Asp	Gly	Lys	Glu	Ile	Pro	Val	Ser	Arg	Val	Glu
	130					135					140				
Lys	Ala	Asp	Pro	Thr	Asp	Ile	Asp	Val	Thr	Asn	Tyr	Val	Arg	Ile	Val
145					150					155					160
Leu	Ser	Glu	Ser	Leu	Lys	Glu	Glu	Asp	Leu	Arg	Lys	Asp	Val	Glu	Leu
			165					170						175	
Ile	Ile	Glu	Gly	Tyr	Lys	Pro	Ala	Arg	Val	Ile	Met	Met	Glu	Ile	Leu
			180					185					190		
Asp	Asp	Tyr	Tyr	Tyr	Asp	Gly	Glu	Leu	Gly	Ala	Val	Tyr	Ser	Pro	Glu
	195						200					205			
Lys	Thr	Ile	Phe	Arg	Val	Trp	Ser	Pro	Val	Ser	Lys	Trp	Val	Lys	Val
	210					215					220				
Leu	Leu	Phe	Lys	Asn	Gly	Glu	Asp	Thr	Glu	Pro	Tyr	Gln	Val	Val	Asn
225					230					235					240
Met	Glu	Tyr	Lys	Gly	Asn	Gly	Val	Trp	Glu	Ala	Val	Val	Glu	Gly	Asp
			245					250						255	
Leu	Asp	Gly	Val	Phe	Tyr	Leu	Tyr	Gln	Leu	Glu	Asn	Tyr	Gly	Lys	Ile
		260						265					270		
Arg	Thr	Thr	Val	Asp	Pro	Tyr	Ser	Lys	Ala	Val	Tyr	Ala	Asn	Asn	Gln
	275						280					285			
Glu	Ser	Ala	Val	Val	Asn	Leu	Ala	Arg	Thr	Asn	Pro	Glu	Gly	Trp	Glu
	290					295					300				
Asn	Asp	Arg	Gly	Pro	Lys	Ile	Glu	Gly	Tyr	Glu	Asp	Ala	Ile	Ile	Tyr
305					310					315					320
Glu	Ile	His	Ile	Ala	Asp	Ile	Thr	Gly	Leu	Glu	Asn	Ser	Gly	Val	Lys
			325					330						335	
Asn	Lys	Gly	Leu	Tyr	Leu	Gly	Leu	Thr	Glu	Glu	Asn	Thr	Lys	Ala	Pro
		340						345					350		
Gly	Gly	Val	Thr	Thr	Gly	Leu	Ser	His	Leu	Val	Glu	Leu	Gly	Val	Thr
		355				360						365			
His	Val	His	Ile	Leu	Pro	Phe	Phe	Asp	Phe	Tyr	Thr	Gly	Asp	Glu	Leu
	370					375					380				
Asp	Lys	Asp	Phe	Glu	Lys	Tyr	Tyr	Asn	Trp	Gly	Tyr	Asp	Pro	Tyr	Leu
385					390					395					400
Phe	Met	Val	Pro	Glu	Gly	Arg	Tyr	Ser	Thr	Asp	Pro	Lys	Asn	Pro	His
			405					410						415	
Thr	Arg	Ile	Arg	Glu	Val	Lys	Glu	Met	Val	Lys	Ala	Leu	His	Lys	His
		420					425						430		
Gly	Ile	Gly	Val	Ile	Met	Asp	Met	Val	Phe	Pro	His	Thr	Tyr	Gly	Ile
	435						440					445			
Gly	Glu	Leu	Ser	Ala	Phe	Asp	Gln	Thr	Val	Pro	Tyr	Tyr	Phe	Tyr	Arg
	450					455					460				
Ile	Asp	Lys	Thr	Gly	Ala	Tyr	Leu	Asn	Glu	Ser	Gly	Cys	Gly	Asn	Val
465					470					475					480
Ile	Ala	Ser	Glu	Arg	Pro	Met	Met	Arg	Lys	Phe	Ile	Val	Asp	Thr	Val
			485					490						495	
Thr	Tyr	Trp	Val	Lys	Glu	Tyr	His	Ile	Asp	Gly	Phe	Arg	Phe	Asp	Gln

```

      500      505      510
Met Gly Leu Ile Asp Lys Lys Thr Met Leu Glu Val Glu Arg Ala Leu
      515      520      525
His Lys Ile Asp Pro Thr Ile Ile Leu Tyr Gly Glu Pro Trp Gly Gly
      530      535      540
Trp Gly Ala Pro Ile Arg Phe Gly Lys Ser Asp Val Ala Gly Thr His
545      550      555      560
Val Ala Ala Phe Asn Asp Glu Phe Arg Asp Ala Ile Arg Gly Ser Val
      565      570      575
Phe Asn Pro Ser Val Lys Gly Phe Val Met Gly Gly Tyr Gly Lys Glu
      580      585      590
Thr Lys Ile Lys Arg Gly Val Val Gly Ser Ile Asn Tyr Asp Gly Lys
      595      600      605
Leu Ile Lys Ser Phe Ala Leu Asp Pro Glu Glu Thr Ile Asn Tyr Ala
      610      615      620
Ala Cys His Asp Asn His Thr Leu Trp Asp Lys Asn Tyr Leu Ala Ala
625      630      635      640
Lys Ala Asp Lys Lys Lys Glu Trp Thr Glu Glu Glu Leu Lys Asn Ala
      645      650      655
Gln Lys Leu Ala Gly Ala Ile Leu Leu Thr Ser Gln Gly Val Pro Phe
      660      665      670
Leu His Gly Gly Gln Asp Phe Cys Arg Thr Thr Asn Phe Asn Asp Asn
      675      680      685
Ser Tyr Asn Ala Pro Ile Ser Ile Asn Gly Phe Asp Tyr Glu Arg Lys
      690      695      700
Leu Gln Phe Ile Asp Val Phe Asn Tyr His Lys Gly Leu Ile Lys Leu
705      710      715      720
Arg Lys Glu His Pro Ala Phe Arg Leu Lys Asn Ala Glu Glu Ile Lys
      725      730      735
Lys His Leu Glu Phe Leu Pro Gly Gly Arg Arg Ile Val Ala Phe Met
      740      745      750
Leu Lys Asp His Ala Gly Gly Asp Pro Trp Lys Asp Ile Val Val Ile
      755      760      765
Tyr Asn Gly Asn Leu Glu Lys Thr Thr Tyr Lys Leu Pro Glu Gly Lys
      770      775      780
Trp Asn Val Val Val Asn Ser Gln Lys Ala Gly Thr Glu Val Ile Glu
785      790      795      800
Thr Val Glu Gly Thr Ile Glu Leu Asp Pro Leu Ser Ala Tyr Val Leu
      805      810      815
Tyr Arg Glu Ser Glu Lys Asp Glu Leu
      820      825

```

&lt;210&gt; 25

&lt;211&gt; 2478

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic

&lt;400&gt; 25

atgagggtgt tgctcgttgc cctcgccttc ctggctctcg ctgcgagcgc caccagcgct 60  
ggccactggt acaagcacca gcgcgcctac cagttcaccg gcgaggacga cttcgggaag 120

```

gtggccgtgg tgaagctccc gatggacctc accaaggtgg gcatcatcgt gcgcctcaac 180
gagtggcagg cgaaggacgt ggccaaggac cgcttcatcg agatcaagga cggcaaggcc 240
gaggtgtgga tactccaggg cgtggaggag atcttctacg agaagccgga cacctccccg 300
cgcattctct tcgcccaggc ccgctccaac aaggtgatcg aggcttctct caccaaccgg 360
gtggacacca agaagaagga gctgttcaag gtgaccgtcg acggcaagga gatcccgggtg 420
tcccgcgtgg agaaggccga cccgaccgac atcgacgtga ccaactacgt gcgcatcgtg 480
ctctccgagt ccctcaagga ggaggacctc cgcaaggacg tggagctgat catcgagggc 540
tacaagccgg cccgcgtgat catgatggag atcctcgacg actactacta cgacggcgag 600
ctgggggagg tgtactcccc ggagaagacc atcttccgcg tgtggtcccc ggtgtccaag 660
tgggtgaagg tgctcctctt caagaacggc gaggacaccg agccgtacca ggtggtgaac 720
atggagtaca agggcaacgg cgtgtgggag gccgtggtgg agggcgacct cgacggcggtg 780
ttctacctct accagctgga gaactacggc aagatccgca ccaccgtgga cccgtactcc 840
aaggccgtgt acgccaacaa ccaggagtct gcagtgggtg acctcgcccc caccaaccgg 900
gagggctggg agaacgaccg cggcccgaag atcgagggtc acgaggacgc catcatctac 960
gagatccaca tcgccgacat caccggcctg gagaactccg gcgtgaagaa caagggcctc 1020
tacctcggcc tcaccgagga gaacaccaag gccccgggcg gcgtgaccac cggcctctcc 1080
cacctcgtgg agctgggctg gaccacgtg cacatcctcc cgttcttcga cttctacacc 1140
ggcgacgagc tggacaagga cttcgagaag tactacaact ggggctacga cccgtacctc 1200
ttcatggtgc cggaggggcg ctactccacc gaccgaaga acccgcacac ccgaattcgc 1260
gaggtgaagg agatggtgaa ggccctccac aagcacggca tcggcggtgat catggacatg 1320
gtgttccccg acacctacgg catcggcgag ctgtccgcct tcgaccagac cgtgccgtac 1380
tacttctacc gcatcgacaa gaccggcgcc tacctcaacg agtccggctg cggcaacgtg 1440
atcgctccg agcgcccgat gatgcgcaag ttcacgtgag acaccgtgac ctactgggtg 1500
aaggagtacc acatcgacgg cttccgcttc gaccagatgg gcctcatcga caagaagacc 1560
atgctggagg tggagcgcg cctccacaag atcgaccgga ccatcatcct ctacggcgag 1620
ccgtggggcg gctggggggc cccgatccgc ttcggcaagt ccgacgtggc cggcaccac 1680
gtggccgcct tcaacgacga gttccgcgac gccatccgcg gctccgtgtt caaccgcgtc 1740
gtgaagggct tcgtgatggg cggctacggc aaggagacca agatcaagcg cggcggtggtg 1800
ggctccatca actacgacgg caagctcatc aagtccttcg ccctcgacct ggaggagacc 1860
atcaactacg ccgcctgcca cgacaaccac accctctggg acaagaacta cctcgccgcc 1920
aaggccgaca agaagaagga gtggaccgag gaggagctga agaacgcccc gaagctcgcc 1980
ggcgccatcc tcctcactag tcaggcggtg ccgttcctcc acggcgggca ggacttctgc 2040
cgcaccacca acttcaacga caactcctac aacgccccga tctccatcaa cggcttcgac 2100
tacgagcgca agctccagtt catcgacgtg ttcaactacc acaagggcct catcaagctc 2160
cgcaaggagc accgggcctt ccgcctcaag aacggcgagg agatcaagaa gcacctggag 2220
ttcttcccgg gcggcgccg catcgtggcc ttcattgctca aggaccacgc cggcggcgac 2280
ccgtggaagg acatcgtggt gatctacaac ggcaacctgg agaagaccac ctacaagctc 2340
ccggagggca agtggaaagt ggtggtgaac tcccagaagg ccggcaccga ggtgatcgag 2400
accgtggagg gcaccatcga gctggaccgg ctctccgcct acgtgctcta ccgagagtcc 2460
gagaaggacg agctgtga 2478

```

<210> 26

<211> 718

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic

<400> 26

```

Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ser
1           5           10           15
Ala Thr Ser Met Glu Thr Ile Lys Ile Tyr Glu Asn Lys Gly Val Tyr
20           25           30

```

Lys Val Val Ile Gly Glu Pro Phe Pro Pro Ile Glu Phe Pro Leu Glu  
 35 40 45  
 Gln Lys Ile Ser Ser Asn Lys Ser Leu Ser Glu Leu Gly Leu Thr Ile  
 50 55 60  
 Val Gln Gln Gly Asn Lys Val Ile Val Glu Lys Ser Leu Asp Leu Lys  
 65 70 75 80  
 Glu His Ile Ile Gly Leu Gly Glu Lys Ala Phe Glu Leu Asp Arg Lys  
 85 90 95  
 Arg Lys Arg Tyr Val Met Tyr Asn Val Asp Ala Gly Ala Tyr Lys Lys  
 100 105 110  
 Tyr Gln Asp Pro Leu Tyr Val Ser Ile Pro Leu Phe Ile Ser Val Lys  
 115 120 125  
 Asp Gly Val Ala Thr Gly Tyr Phe Phe Asn Ser Ala Ser Lys Val Ile  
 130 135 140  
 Phe Asp Val Gly Leu Glu Glu Tyr Asp Lys Val Ile Val Thr Ile Pro  
 145 150 155 160  
 Glu Asp Ser Val Glu Phe Tyr Val Ile Glu Gly Pro Arg Ile Glu Asp  
 165 170 175  
 Val Leu Glu Lys Tyr Thr Glu Leu Thr Gly Lys Pro Phe Leu Pro Pro  
 180 185 190  
 Met Trp Ala Phe Gly Tyr Met Ile Ser Arg Tyr Ser Tyr Tyr Pro Gln  
 195 200 205  
 Asp Lys Val Val Glu Leu Val Asp Ile Met Gln Lys Glu Gly Phe Arg  
 210 215 220  
 Val Ala Gly Val Phe Leu Asp Ile His Tyr Met Asp Ser Tyr Lys Leu  
 225 230 235 240  
 Phe Thr Trp His Pro Tyr Arg Phe Pro Glu Pro Lys Lys Leu Ile Asp  
 245 250 255  
 Glu Leu His Lys Arg Asn Val Lys Leu Ile Thr Ile Val Asp His Gly  
 260 265 270  
 Ile Arg Val Asp Gln Asn Tyr Ser Pro Phe Leu Ser Gly Met Gly Lys  
 275 280 285  
 Phe Cys Glu Ile Glu Ser Gly Glu Leu Phe Val Gly Lys Met Trp Pro  
 290 295 300  
 Gly Thr Thr Val Tyr Pro Asp Phe Phe Arg Glu Asp Thr Arg Glu Trp  
 305 310 315 320  
 Trp Ala Gly Leu Ile Ser Glu Trp Leu Ser Gln Gly Val Asp Gly Ile  
 325 330 335  
 Trp Leu Asp Met Asn Glu Pro Thr Asp Phe Ser Arg Ala Ile Glu Ile  
 340 345 350  
 Arg Asp Val Leu Ser Ser Leu Pro Val Gln Phe Arg Asp Asp Arg Leu  
 355 360 365  
 Val Thr Thr Phe Pro Asp Asn Val Val His Tyr Leu Arg Gly Lys Arg  
 370 375 380  
 Val Lys His Glu Lys Val Arg Asn Ala Tyr Pro Leu Tyr Glu Ala Met  
 385 390 395 400  
 Ala Thr Phe Lys Gly Phe Arg Thr Ser His Arg Asn Glu Ile Phe Ile  
 405 410 415  
 Leu Ser Arg Ala Gly Tyr Ala Gly Ile Gln Arg Tyr Ala Phe Ile Trp  
 420 425 430  
 Thr Gly Asp Asn Thr Pro Ser Trp Asp Asp Leu Lys Leu Gln Leu Gln  
 435 440 445  
 Leu Val Leu Gly Leu Ser Ile Ser Gly Val Pro Phe Val Gly Cys Asp  
 450 455 460

```

Ile Gly Gly Phe Gln Gly Arg Asn Phe Ala Glu Ile Asp Asn Ser Met
465          470          475          480
Asp Leu Leu Val Lys Tyr Tyr Ala Leu Ala Leu Phe Phe Pro Phe Tyr
          485          490          495
Arg Ser His Lys Ala Thr Asp Gly Ile Asp Thr Glu Pro Val Phe Leu
          500          505          510
Pro Asp Tyr Tyr Lys Glu Lys Val Lys Glu Ile Val Glu Leu Arg Tyr
          515          520          525
Lys Phe Leu Pro Tyr Ile Tyr Ser Leu Ala Leu Glu Ala Ser Glu Lys
          530          535          540
Gly His Pro Val Ile Arg Pro Leu Phe Tyr Glu Phe Gln Asp Asp Asp
545          550          555          560
Asp Met Tyr Arg Ile Glu Asp Glu Tyr Met Val Gly Lys Tyr Leu Leu
          565          570          575
Tyr Ala Pro Ile Val Ser Lys Glu Glu Ser Arg Leu Val Thr Leu Pro
          580          585          590
Arg Gly Lys Trp Tyr Asn Tyr Trp Asn Gly Glu Ile Ile Asn Gly Lys
          595          600          605
Ser Val Val Lys Ser Thr His Glu Leu Pro Ile Tyr Leu Arg Glu Gly
          610          615          620
Ser Ile Ile Pro Leu Glu Gly Asp Glu Leu Ile Val Tyr Gly Glu Thr
625          630          635          640
Ser Phe Lys Arg Tyr Asp Asn Ala Glu Ile Thr Ser Ser Ser Asn Glu
          645          650          655
Ile Lys Phe Ser Arg Glu Ile Tyr Val Ser Lys Leu Thr Ile Thr Ser
          660          665          670
Glu Lys Pro Val Ser Lys Ile Ile Val Asp Asp Ser Lys Glu Ile Gln
          675          680          685
Val Glu Lys Thr Met Gln Asn Thr Tyr Val Ala Lys Ile Asn Gln Lys
          690          695          700
Ile Arg Gly Lys Ile Asn Leu Glu Ser Glu Lys Asp Glu Leu
705          710          715

```

&lt;210&gt; 27

&lt;211&gt; 712

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic

&lt;400&gt; 27

```

Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser
1          5          10          15
Ala Thr Ser Met Glu Thr Ile Lys Ile Tyr Glu Asn Lys Gly Val Tyr
          20          25          30
Lys Val Val Ile Gly Glu Pro Phe Pro Pro Ile Glu Phe Pro Leu Glu
          35          40          45
Gln Lys Ile Ser Ser Asn Lys Ser Leu Ser Glu Leu Gly Leu Thr Ile
          50          55          60
Val Gln Gln Gly Asn Lys Val Ile Val Glu Lys Ser Leu Asp Leu Lys
65          70          75          80
Glu His Ile Ile Gly Leu Gly Glu Lys Ala Phe Glu Leu Asp Arg Lys

```



197

```

      515              520              525 -
Lys Phe Leu Pro Tyr Ile Tyr Ser Leu Ala Leu Glu Ala Ser Glu Lys
  530              535              540
Gly His Pro Val Ile Arg Pro Leu Phe Tyr Glu Phe Gln Asp Asp Asp
  545              550              555              560
Asp Met Tyr Arg Ile Glu Asp Glu Tyr Met Val Gly Lys Tyr Leu Leu
      565              570              575
Tyr Ala Pro Ile Val Ser Lys Glu Glu Ser Arg Leu Val Thr Leu Pro
      580              585              590
Arg Gly Lys Trp Tyr Asn Tyr Trp Asn Gly Glu Ile Ile Asn Gly Lys
      595              600              605
Ser Val Val Lys Ser Thr His Glu Leu Pro Ile Tyr Leu Arg Glu Gly
      610              615              620
Ser Ile Ile Pro Leu Glu Gly Asp Glu Leu Ile Val Tyr Gly Glu Thr
  625              630              635              640
Ser Phe Lys Arg Tyr Asp Asn Ala Glu Ile Thr Ser Ser Ser Asn Glu
      645              650              655
Ile Lys Phe Ser Arg Glu Ile Tyr Val Ser Lys Leu Thr Ile Thr Ser
      660              665              670
Glu Lys Pro Val Ser Lys Ile Ile Val Asp Asp Ser Lys Glu Ile Gln
      675              680              685
Val Glu Lys Thr Met Gln Asn Thr Tyr Val Ala Lys Ile Asn Gln Lys
      690              695              700
Ile Arg Gly Lys Ile Asn Leu Glu
  705              710

```

&lt;210&gt; 28

&lt;211&gt; 469

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic

&lt;400&gt; 28

```

Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Ala Ser
  1              5              10              15
Ala Thr Ser Met Ala Glu Phe Phe Pro Glu Ile Pro Lys Ile Gln Phe
      20              25              30
Glu Gly Lys Glu Ser Thr Asn Pro Leu Ala Phe Arg Phe Tyr Asp Pro
      35              40              45
Asn Glu Val Ile Asp Gly Lys Pro Leu Lys Asp His Leu Lys Phe Ser
      50              55              60
Val Ala Phe Trp His Thr Phe Val Asn Glu Gly Arg Asp Pro Phe Gly
  65              70              75              80
Asp Pro Thr Ala Glu Arg Pro Trp Asn Arg Phe Ser Asp Pro Met Asp
      85              90              95
Lys Ala Phe Ala Arg Val Asp Ala Leu Phe Glu Phe Cys Glu Lys Leu
      100              105              110
Asn Ile Glu Tyr Phe Cys Phe His Asp Arg Asp Ile Ala Pro Glu Gly
      115              120              125
Lys Thr Leu Arg Glu Thr Asn Lys Ile Leu Asp Lys Val Val Glu Arg
  130              135              140

```

```

Ile Lys Glu Arg Met Lys Asp Ser Asn Val Lys Leu Leu Trp Gly Thr
145          150          155          160
Ala Asn Leu Phe Ser His Pro Arg Tyr Met His Gly Ala Ala Thr Thr
          165          170          175
Cys Ser Ala Asp Val Phe Ala Tyr Ala Ala Ala Gln Val Lys Lys Ala
          180          185          190
Leu Glu Ile Thr Lys Glu Leu Gly Gly Glu Gly Tyr Val Phe Trp Gly
          195          200          205
Gly Arg Glu Gly Tyr Glu Thr Leu Leu Asn Thr Asp Leu Gly Leu Glu
          210          215          220
Leu Glu Asn Leu Ala Arg Phe Leu Arg Met Ala Val Glu Tyr Ala Lys
          225          230          235          240
Lys Ile Gly Phe Thr Gly Gln Phe Leu Ile Glu Pro Lys Pro Lys Glu
          245          250          255
Pro Thr Lys His Gln Tyr Asp Phe Asp Val Ala Thr Ala Tyr Ala Phe
          260          265          270
Leu Lys Asn His Gly Leu Asp Glu Tyr Phe Lys Phe Asn Ile Glu Ala
          275          280          285
Asn His Ala Thr Leu Ala Gly His Thr Phe Gln His Glu Leu Arg Met
          290          295          300
Ala Arg Ile Leu Gly Lys Leu Gly Ser Ile Asp Ala Asn Gln Gly Asp
          305          310          315          320
Leu Leu Leu Gly Trp Asp Thr Asp Gln Phe Pro Thr Asn Ile Tyr Asp
          325          330          335
Thr Thr Leu Ala Met Tyr Glu Val Ile Lys Ala Gly Gly Phe Thr Lys
          340          345          350
Gly Gly Leu Asn Phe Asp Ala Lys Val Arg Arg Ala Ser Tyr Lys Val
          355          360          365
Glu Asp Leu Phe Ile Gly His Ile Ala Gly Met Asp Thr Phe Ala Leu
          370          375          380
Gly Phe Lys Ile Ala Tyr Lys Leu Ala Lys Asp Gly Val Phe Asp Lys
          385          390          395          400
Phe Ile Glu Glu Lys Tyr Arg Ser Phe Lys Glu Gly Ile Gly Lys Glu
          405          410          415
Ile Val Glu Gly Lys Thr Asp Phe Glu Lys Leu Glu Glu Tyr Ile Ile
          420          425          430
Asp Lys Glu Asp Ile Glu Leu Pro Ser Gly Lys Gln Glu Tyr Leu Glu
          435          440          445
Ser Leu Leu Asn Ser Tyr Ile Val Lys Thr Ile Ala Glu Leu Arg Ser
          450          455          460
Glu Lys Asp Glu Leu
465

```

&lt;210&gt; 29

&lt;211&gt; 469

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic

&lt;400&gt; 29

Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser

1	5	10	15
Ala Thr Ser Met	Ala Glu Phe Phe	Pro Glu Ile Pro	Lys Val Gln Phe
20	25	30	
Glu Gly Lys Glu	Ser Thr Asn Pro	Leu Ala Phe Lys	Phe Tyr Asp Pro
35	40	45	
Glu Glu Ile Ile	Asp Gly Lys Pro	Leu Lys Asp His	Leu Lys Phe Ser
50	55	60	
Val Ala Phe Trp	His Thr Phe Val	Asn Glu Gly Arg	Asp Pro Phe Gly
65	70	75	80
Asp Pro Thr Ala	Asp Arg Pro Trp	Asn Arg Tyr Thr	Asp Pro Met Asp
85	90	95	
Lys Ala Phe Ala	Arg Val Asp Ala	Leu Phe Glu Phe	Cys Glu Lys Leu
100	105	110	
Asn Ile Glu Tyr	Phe Cys Phe His	Asp Arg Asp Ile	Ala Pro Glu Gly
115	120	125	
Lys Thr Leu Arg	Glu Thr Asn Lys	Ile Leu Asp Lys	Val Val Glu Arg
130	135	140	
Ile Lys Glu Arg	Met Lys Asp Ser	Asn Val Lys Leu	Leu Trp Gly Thr
145	150	155	160
Ala Asn Leu Phe	Ser His Pro Arg	Tyr Met His Gly	Ala Ala Thr Thr
165	170	175	
Cys Ser Ala Asp	Val Phe Ala Tyr	Ala Ala Gln Val	Lys Lys Ala
180	185	190	
Leu Glu Ile Thr	Lys Glu Leu Gly	Gly Glu Tyr Val	Phe Trp Gly
195	200	205	
Gly Arg Glu Gly	Tyr Glu Thr Leu	Leu Asn Thr Asp	Leu Gly Phe Glu
210	215	220	
Leu Glu Asn Leu	Ala Arg Phe Leu	Arg Met Ala Val	Asp Tyr Ala Lys
225	230	235	240
Arg Ile Gly Phe	Thr Gly Gln Phe	Leu Ile Glu Pro	Lys Pro Lys Glu
245	250	255	
Pro Thr Lys His	Gln Tyr Asp Phe	Asp Val Ala Thr	Ala Tyr Ala Phe
260	265	270	
Leu Lys Ser His	Gly Leu Asp Glu	Tyr Phe Lys Phe	Asn Ile Glu Ala
275	280	285	
Asn His Ala Thr	Leu Ala Gly His	Thr Phe Gln His	Glu Leu Arg Met
290	295	300	
Ala Arg Ile Leu	Gly Lys Leu Gly	Ser Ile Asp Ala	Asn Gln Gly Asp
305	310	315	320
Leu Leu Leu Gly	Trp Asp Thr Asp	Gln Phe Pro Thr	Asn Val Tyr Asp
325	330	335	
Thr Thr Leu Ala	Met Tyr Glu Val	Ile Lys Ala Gly	Gly Phe Thr Lys
340	345	350	
Gly Gly Leu Asn	Phe Asp Ala Lys	Val Arg Arg Ala	Ser Tyr Lys Val
355	360	365	
Glu Asp Leu Phe	Ile Gly His Ile	Ala Gly Met Asp	Thr Phe Ala Leu
370	375	380	
Gly Phe Lys Val	Ala Tyr Lys Leu	Val Lys Asp Gly	Val Leu Asp Lys
385	390	395	400
Phe Ile Glu Glu	Lys Tyr Arg Ser	Phe Arg Glu Gly	Ile Gly Arg Asp
405	410	415	
Ile Val Glu Gly	Lys Val Asp Phe	Glu Lys Leu Glu	Glu Tyr Ile Ile
420	425	430	
Asp Lys Glu Thr	Ile Glu Leu Pro	Ser Gly Lys Gln	Glu Tyr Leu Glu

435                      440                      445  
 Ser Leu Ile Asn Ser Tyr Ile Val Lys Thr Ile Leu Glu Leu Arg Ser  
 450                      455                      460  
 Glu Lys Asp Glu Leu  
 465

<210> 30  
 <211> 463  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 30  
 Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser  
 1                      5                      10                      15  
 Ala Thr Ser Met Ala Glu Phe Phe Pro Glu Ile Pro Lys Val Gln Phe  
 20                      25                      30  
 Glu Gly Lys Glu Ser Thr Asn Pro Leu Ala Phe Lys Phe Tyr Asp Pro  
 35                      40                      45  
 Glu Glu Ile Ile Asp Gly Lys Pro Leu Lys Asp His Leu Lys Phe Ser  
 50                      55                      60  
 Val Ala Phe Trp His Thr Phe Val Asn Glu Gly Arg Asp Pro Phe Gly  
 65                      70                      75                      80  
 Asp Pro Thr Ala Asp Arg Pro Trp Asn Arg Tyr Thr Asp Pro Met Asp  
 85                      90                      95  
 Lys Ala Phe Ala Arg Val Asp Ala Leu Phe Glu Phe Cys Glu Lys Leu  
 100                      105                      110  
 Asn Ile Glu Tyr Phe Cys Phe His Asp Arg Asp Ile Ala Pro Glu Gly  
 115                      120                      125  
 Lys Thr Leu Arg Glu Thr Asn Lys Ile Leu Asp Lys Val Val Glu Arg  
 130                      135                      140  
 Ile Lys Glu Arg Met Lys Asp Ser Asn Val Lys Leu Leu Trp Gly Thr  
 145                      150                      155                      160  
 Ala Asn Leu Phe Ser His Pro Arg Tyr Met His Gly Ala Ala Thr Thr  
 165                      170                      175  
 Cys Ser Ala Asp Val Phe Ala Tyr Ala Ala Ala Gln Val Lys Lys Ala  
 180                      185                      190  
 Leu Glu Ile Thr Lys Glu Leu Gly Gly Glu Gly Tyr Val Phe Trp Gly  
 195                      200                      205  
 Gly Arg Glu Gly Tyr Glu Thr Leu Leu Asn Thr Asp Leu Gly Phe Glu  
 210                      215                      220  
 Leu Glu Asn Leu Ala Arg Phe Leu Arg Met Ala Val Asp Tyr Ala Lys  
 225                      230                      235                      240  
 Arg Ile Gly Phe Thr Gly Gln Phe Leu Ile Glu Pro Lys Pro Lys Glu  
 245                      250                      255  
 Pro Thr Lys His Gln Tyr Asp Phe Asp Val Ala Thr Ala Tyr Ala Phe  
 260                      265                      270  
 Leu Lys Ser His Gly Leu Asp Glu Tyr Phe Lys Phe Asn Ile Glu Ala  
 275                      280                      285  
 Asn His Ala Thr Leu Ala Gly His Thr Phe Gln His Glu Leu Arg Met  
 290                      295                      300

Ala Arg Ile Leu Gly Lys Leu Gly Ser Ile Asp Ala Asn Gln Gly Asp  
 305 310 315 320  
 Leu Leu Leu Gly Trp Asp Thr Asp Gln Phe Pro Thr Asn Val Tyr Asp  
 325 330 335  
 Thr Thr Leu Ala Met Tyr Glu Val Ile Lys Ala Gly Gly Phe Thr Lys  
 340 345 350  
 Gly Gly Leu Asn Phe Asp Ala Lys Val Arg Arg Ala Ser Tyr Lys Val  
 355 360 365  
 Glu Asp Leu Phe Ile Gly His Ile Ala Gly Met Asp Thr Phe Ala Leu  
 370 375 380  
 Gly Phe Lys Val Ala Tyr Lys Leu Val Lys Asp Gly Val Leu Asp Lys  
 385 390 395 400  
 Phe Ile Glu Glu Lys Tyr Arg Ser Phe Arg Glu Gly Ile Gly Arg Asp  
 405 410 415  
 Ile Val Glu Gly Lys Val Asp Phe Glu Lys Leu Glu Glu Tyr Ile Ile  
 420 425 430  
 Asp Lys Glu Thr Ile Glu Leu Pro Ser Gly Lys Gln Glu Tyr Leu Glu  
 435 440 445  
 Ser Leu Ile Asn Ser Tyr Ile Val Lys Thr Ile Leu Glu Leu Arg  
 450 455 460

<210> 31  
 <211> 25  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 31  
 Met Gly Lys Asn Gly Asn Leu Cys Cys Phe Ser Leu Leu Leu Leu Leu  
 1 5 10 15  
 Leu Ala Gly Leu Ala Ser Gly His Gln  
 20 25

<210> 32  
 <211> 30  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 32  
 Met Gly Phe Val Leu Phe Ser Gln Leu Pro Ser Phe Leu Leu Val Ser  
 1 5 10 15  
 Thr Leu Leu Leu Phe Leu Val Ile Ser His Ser Cys Arg Ala  
 20 25 30

<210> 33  
 <211> 460

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic

&lt;400&gt; 33

```

Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser
 1          5          10          15
Ala Thr Ser Ala Lys Tyr Leu Glu Leu Glu Gly Gly Val Ile Met
 20          25          30
Gln Ala Phe Tyr Trp Asp Val Pro Ser Gly Gly Ile Trp Trp Asp Thr
 35          40          45
Ile Arg Gln Lys Ile Pro Glu Trp Tyr Asp Ala Gly Ile Ser Ala Ile
 50          55          60
Trp Ile Pro Pro Ala Ser Lys Gly Met Ser Gly Gly Tyr Ser Met Gly
 65          70          75          80
Tyr Asp Pro Tyr Asp Tyr Phe Asp Leu Gly Glu Tyr Tyr Gln Lys Gly
 85          90          95
Thr Val Glu Thr Arg Phe Gly Ser Lys Gln Glu Leu Ile Asn Met Ile
100          105          110
Asn Thr Ala His Ala Tyr Gly Ile Lys Val Ile Ala Asp Ile Val Ile
115          120          125
Asn His Arg Ala Gly Gly Asp Leu Glu Trp Asn Pro Phe Val Gly Asp
130          135          140
Tyr Thr Trp Thr Asp Phe Ser Lys Val Ala Ser Gly Lys Tyr Thr Ala
145          150          155          160
Asn Tyr Leu Asp Phe His Pro Asn Glu Leu His Ala Gly Asp Ser Gly
165          170          175
Thr Phe Gly Gly Tyr Pro Asp Ile Cys His Asp Lys Ser Trp Asp Gln
180          185          190
Tyr Trp Leu Trp Ala Ser Gln Glu Ser Tyr Ala Ala Tyr Leu Arg Ser
195          200          205
Ile Gly Ile Asp Ala Trp Arg Phe Asp Tyr Val Lys Gly Tyr Gly Ala
210          215          220
Trp Val Val Lys Asp Trp Leu Asn Trp Trp Gly Gly Trp Ala Val Gly
225          230          235          240
Glu Tyr Trp Asp Thr Asn Val Asp Ala Leu Leu Asn Trp Ala Tyr Ser
245          250          255
Ser Gly Ala Lys Val Phe Asp Phe Pro Leu Tyr Tyr Lys Met Asp Ala
260          265          270
Ala Phe Asp Asn Lys Asn Ile Pro Ala Leu Val Glu Ala Leu Lys Asn
275          280          285
Gly Gly Thr Val Val Ser Arg Asp Pro Phe Lys Ala Val Thr Phe Val
290          295          300
Ala Asn His Asp Thr Asp Ile Ile Trp Asn Lys Tyr Pro Ala Tyr Ala
305          310          315          320
Phe Ile Leu Thr Tyr Glu Gly Gln Pro Thr Ile Phe Tyr Arg Asp Tyr
325          330          335
Glu Glu Trp Leu Asn Lys Asp Lys Leu Lys Asn Leu Ile Trp Ile His
340          345          350
Asp Asn Leu Ala Gly Gly Ser Thr Ser Ile Val Tyr Tyr Asp Ser Asp
355          360          365
Glu Met Ile Phe Val Arg Asn Gly Tyr Gly Ser Lys Pro Gly Leu Ile

```

```

      370      375      380
Thr Tyr Ile Asn Leu Gly Ser Ser Lys Val Gly Arg Trp Val Tyr Val
385      390      395      400
Pro Lys Phe Ala Gly Ala Cys Ile His Glu Tyr Thr Gly Asn Leu Gly
      405      410      415
Gly Trp Val Asp Lys Tyr Val Tyr Ser Ser Gly Trp Val Tyr Leu Glu
      420      425      430
Ala Pro Ala Tyr Asp Pro Ala Asn Gly Gln Tyr Gly Tyr Ser Val Trp
      435      440      445
Ser Tyr Cys Gly Val Gly Ser Glu Lys Asp Glu Leu
      450      455      460

```

<210> 34  
 <211> 825  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> synthetic

```

<400> 34
Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser
 1      5      10      15
Ala Thr Ser Ala Gly His Trp Tyr Lys His Gln Arg Ala Tyr Gln Phe
      20      25      30
Thr Gly Glu Asp Asp Phe Gly Lys Val Ala Val Val Lys Leu Pro Met
      35      40      45
Asp Leu Thr Lys Val Gly Ile Ile Val Arg Leu Asn Glu Trp Gln Ala
      50      55      60
Lys Asp Val Ala Lys Asp Arg Phe Ile Glu Ile Lys Asp Gly Lys Ala
      65      70      75      80
Glu Val Trp Ile Leu Gln Gly Val Glu Glu Ile Phe Tyr Glu Lys Pro
      85      90      95
Asp Thr Ser Pro Arg Ile Phe Phe Ala Gln Ala Arg Ser Asn Lys Val
      100      105      110
Ile Glu Ala Phe Leu Thr Asn Pro Val Asp Thr Lys Lys Lys Glu Leu
      115      120      125
Phe Lys Val Thr Val Asp Gly Lys Glu Ile Pro Val Ser Arg Val Glu
      130      135      140
Lys Ala Asp Pro Thr Asp Ile Asp Val Thr Asn Tyr Val Arg Ile Val
      145      150      155      160
Leu Ser Glu Ser Leu Lys Glu Glu Asp Leu Arg Lys Asp Val Glu Leu
      165      170      175
Ile Ile Glu Gly Tyr Lys Pro Ala Arg Val Ile Met Met Glu Ile Leu
      180      185      190
Asp Asp Tyr Tyr Tyr Asp Gly Glu Leu Gly Ala Val Tyr Ser Pro Glu
      195      200      205
Lys Thr Ile Phe Arg Val Trp Ser Pro Val Ser Lys Trp Val Lys Val
      210      215      220
Leu Leu Phe Lys Asn Gly Glu Asp Thr Glu Pro Tyr Gln Val Val Asn
      225      230      235      240
Met Glu Tyr Lys Gly Asn Gly Val Trp Glu Ala Val Val Glu Gly Asp
      245      250      255

```



Leu Asp Gly Val Phe Tyr Leu Tyr Gln Leu Glu Asn Tyr Gly Lys Ile  
 260 265 270  
 Arg Thr Thr Val Asp Pro Tyr Ser Lys Ala Val Tyr Ala Asn Asn Gln  
 275 280 285  
 Glu Ser Ala Val Val Asn Leu Ala Arg Thr Asn Pro Glu Gly Trp Glu  
 290 295 300  
 Asn Asp Arg Gly Pro Lys Ile Glu Gly Tyr Glu Asp Ala Ile Ile Tyr  
 305 310 315 320  
 Glu Ile His Ile Ala Asp Ile Thr Gly Leu Glu Asn Ser Gly Val Lys  
 325 330 335  
 Asn Lys Gly Leu Tyr Leu Gly Leu Thr Glu Glu Asn Thr Lys Ala Pro  
 340 345 350  
 Gly Gly Val Thr Thr Gly Leu Ser His Leu Val Glu Leu Gly Val Thr  
 355 360 365  
 His Val His Ile Leu Pro Phe Phe Asp Phe Tyr Thr Gly Asp Glu Leu  
 370 375 380  
 Asp Lys Asp Phe Glu Lys Tyr Tyr Asn Trp Gly Tyr Asp Pro Tyr Leu  
 385 390 395 400  
 Phe Met Val Pro Glu Gly Arg Tyr Ser Thr Asp Pro Lys Asn Pro His  
 405 410 415  
 Thr Arg Ile Arg Glu Val Lys Glu Met Val Lys Ala Leu His Lys His  
 420 425 430  
 Gly Ile Gly Val Ile Met Asp Met Val Phe Pro His Thr Tyr Gly Ile  
 435 440 445  
 Gly Glu Leu Ser Ala Phe Asp Gln Thr Val Pro Tyr Tyr Phe Tyr Arg  
 450 455 460  
 Ile Asp Lys Thr Gly Ala Tyr Leu Asn Glu Ser Gly Cys Gly Asn Val  
 465 470 475 480  
 Ile Ala Ser Glu Arg Pro Met Met Arg Lys Phe Ile Val Asp Thr Val  
 485 490 495  
 Thr Tyr Trp Val Lys Glu Tyr His Ile Asp Gly Phe Arg Phe Asp Gln  
 500 505 510  
 Met Gly Leu Ile Asp Lys Lys Thr Met Leu Glu Val Glu Arg Ala Leu  
 515 520 525  
 His Lys Ile Asp Pro Thr Ile Ile Leu Tyr Gly Glu Pro Trp Gly Gly  
 530 535 540  
 Trp Gly Ala Pro Ile Arg Phe Gly Lys Ser Asp Val Ala Gly Thr His  
 545 550 555 560  
 Val Ala Ala Phe Asn Asp Glu Phe Arg Asp Ala Ile Arg Gly Ser Val  
 565 570 575  
 Phe Asn Pro Ser Val Lys Gly Phe Val Met Gly Gly Tyr Gly Lys Glu  
 580 585 590  
 Thr Lys Ile Lys Arg Gly Val Val Gly Ser Ile Asn Tyr Asp Gly Lys  
 595 600 605  
 Leu Ile Lys Ser Phe Ala Leu Asp Pro Glu Glu Thr Ile Asn Tyr Ala  
 610 615 620  
 Ala Cys His Asp Asn His Thr Leu Trp Asp Lys Asn Tyr Leu Ala Ala  
 625 630 635 640  
 Lys Ala Asp Lys Lys Lys Glu Trp Thr Glu Glu Glu Leu Lys Asn Ala  
 645 650 655  
 Gln Lys Leu Ala Gly Ala Ile Leu Leu Thr Ser Gln Gly Val Pro Phe  
 660 665 670  
 Leu His Gly Gly Gln Asp Phe Cys Arg Thr Thr Asn Phe Asn Asp Asn  
 675 680 685

```

Ser Tyr Asn Ala Pro Ile Ser Ile Asn Gly Phe Asp Tyr Glu Arg Lys
  690                695                700
Leu Gln Phe Ile Asp Val Phe Asn Tyr His Lys Gly Leu Ile Lys Leu
  705                710                715                720
Arg Lys Glu His Pro Ala Phe Arg Leu Lys Asn Ala Glu Glu Ile Lys
                725                730                735
Lys His Leu Glu Phe Leu Pro Gly Gly Arg Arg Ile Val Ala Phe Met
                740                745                750
Leu Lys Asp His Ala Gly Gly Asp Pro Trp Lys Asp Ile Val Val Ile
                755                760                765
Tyr Asn Gly Asn Leu Glu Lys Thr Thr Tyr Lys Leu Pro Glu Gly Lys
                770                775                780
Trp Asn Val Val Val Asn Ser Gln Lys Ala Gly Thr Glu Val Ile Glu
  785                790                795                800
Thr Val Glu Gly Thr Ile Glu Leu Asp Pro Leu Ser Ala Tyr Val Leu
                805                810                815
Tyr Arg Glu Ser Glu Lys Asp Glu Leu
                820                825

```

<210> 35  
 <211> 460  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> synthetic

```

<400> 35
Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser
  1                5                10                15
Ala Thr Ser Ala Lys Tyr Leu Glu Leu Glu Glu Gly Gly Val Ile Met
                20                25                30
Gln Ala Phe Tyr Trp Asp Val Pro Ser Gly Gly Ile Trp Trp Asp Thr
                35                40                45
Ile Arg Gln Lys Ile Pro Glu Trp Tyr Asp Ala Gly Ile Ser Ala Ile
                50                55                60
Trp Ile Pro Pro Ala Ser Lys Gly Met Ser Gly Gly Tyr Ser Met Gly
  65                70                75                80
Tyr Asp Pro Tyr Asp Tyr Phe Asp Leu Gly Glu Tyr Tyr Gln Lys Gly
                85                90                95
Thr Val Glu Thr Arg Phe Gly Ser Lys Gln Glu Leu Ile Asn Met Ile
                100                105                110
Asn Thr Ala His Ala Tyr Gly Ile Lys Val Ile Ala Asp Ile Val Ile
                115                120                125
Asn His Arg Ala Gly Gly Asp Leu Glu Trp Asn Pro Phe Val Gly Asp
                130                135                140
Tyr Thr Trp Thr Asp Phe Ser Lys Val Ala Ser Gly Lys Tyr Thr Ala
  145                150                155                160
Asn Tyr Leu Asp Phe His Pro Asn Glu Leu His Ala Gly Asp Ser Gly
                165                170                175
Thr Phe Gly Gly Tyr Pro Asp Ile Cys His Asp Lys Ser Trp Asp Gln
                180                185                190
Tyr Trp Leu Trp Ala Ser Gln Glu Ser Tyr Ala Ala Tyr Leu Arg Ser

```

```

      195              200              205
Ile Gly Ile Asp Ala Trp Arg Phe Asp Tyr Val Lys Gly Tyr Gly Ala
  210              215              220
Trp Val Val Lys Asp Trp Leu Asn Trp Trp Gly Gly Trp Ala Val Gly
  225              230              235              240
Glu Tyr Trp Asp Thr Asn Val Asp Ala Leu Leu Asn Trp Ala Tyr Ser
      245              250              255
Ser Gly Ala Lys Val Phe Asp Phe Pro Leu Tyr Tyr Lys Met Asp Ala
      260              265              270
Ala Phe Asp Asn Lys Asn Ile Pro Ala Leu Val Glu Ala Leu Lys Asn
      275              280              285
Gly Gly Thr Val Val Ser Arg Asp Pro Phe Lys Ala Val Thr Phe Val
      290              295              300
Ala Asn His Asp Thr Asp Ile Ile Trp Asn Lys Tyr Pro Ala Tyr Ala
  305              310              315              320
Phe Ile Leu Thr Tyr Glu Gly Gln Pro Thr Ile Phe Tyr Arg Asp Tyr
      325              330              335
Glu Glu Trp Leu Asn Lys Asp Lys Leu Lys Asn Leu Ile Trp Ile His
      340              345              350
Asp Asn Leu Ala Gly Gly Ser Thr Ser Ile Val Tyr Tyr Asp Ser Asp
      355              360              365
Glu Met Ile Phe Val Arg Asn Gly Tyr Gly Ser Lys Pro Gly Leu Ile
      370              375              380
Thr Tyr Ile Asn Leu Gly Ser Ser Lys Val Gly Arg Trp Val Tyr Val
  385              390              395              400
Pro Lys Phe Ala Gly Ala Cys Ile His Glu Tyr Thr Gly Asn Leu Gly
      405              410              415
Gly Trp Val Asp Lys Tyr Val Tyr Ser Ser Gly Trp Val Tyr Leu Glu
      420              425              430
Ala Pro Ala Tyr Asp Pro Ala Asn Gly Gln Tyr Gly Tyr Ser Val Trp
      435              440              445
Ser Tyr Cys Gly Val Gly Ser Glu Lys Asp Glu Leu
      450              455              460

```

<210> 36  
 <211> 718  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> synthetic

```

<400> 36
Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser
  1              5              10              15
Ala Thr Ser Met Glu Thr Ile Lys Ile Tyr Glu Asn Lys Gly Val Tyr
      20              25              30
Lys Val Val Ile Gly Glu Pro Phe Pro Pro Ile Glu Phe Pro Leu Glu
      35              40              45
Gln Lys Ile Ser Ser Asn Lys Ser Leu Ser Glu Leu Gly Leu Thr Ile
      50              55              60
Val Gln Gln Gly Asn Lys Val Ile Val Glu Lys Ser Leu Asp Leu Lys
      65              70              75              80

```

Glu His Ile Ile Gly Leu Gly Glu Lys Ala Phe Glu Leu Asp Arg Lys  
                             85                            90                            95  
 Arg Lys Arg Tyr Val Met Tyr Asn Val Asp Ala Gly Ala Tyr Lys Lys  
                             100                            105                            110  
 Tyr Gln Asp Pro Leu Tyr Val Ser Ile Pro Leu Phe Ile Ser Val Lys  
                             115                            120                            125  
 Asp Gly Val Ala Thr Gly Tyr Phe Phe Asn Ser Ala Ser Lys Val Ile  
                             130                            135                            140  
 Phe Asp Val Gly Leu Glu Glu Tyr Asp Lys Val Ile Val Thr Ile Pro  
 145                            150                            155                            160  
 Glu Asp Ser Val Glu Phe Tyr Val Ile Glu Gly Pro Arg Ile Glu Asp  
                             165                            170                            175  
 Val Leu Glu Lys Tyr Thr Glu Leu Thr Gly Lys Pro Phe Leu Pro Pro  
                             180                            185                            190  
 Met Trp Ala Phe Gly Tyr Met Ile Ser Arg Tyr Ser Tyr Tyr Pro Gln  
                             195                            200                            205  
 Asp Lys Val Val Glu Leu Val Asp Ile Met Gln Lys Glu Gly Phe Arg  
                             210                            215                            220  
 Val Ala Gly Val Phe Leu Asp Ile His Tyr Met Asp Ser Tyr Lys Leu  
 225                            230                            235                            240  
 Phe Thr Trp His Pro Tyr Arg Phe Pro Glu Pro Lys Lys Leu Ile Asp  
                             245                            250                            255  
 Glu Leu His Lys Arg Asn Val Lys Leu Ile Thr Ile Val Asp His Gly  
                             260                            265                            270  
 Ile Arg Val Asp Gln Asn Tyr Ser Pro Phe Leu Ser Gly Met Gly Lys  
                             275                            280                            285  
 Phe Cys Glu Ile Glu Ser Gly Glu Leu Phe Val Gly Lys Met Trp Pro  
                             290                            295                            300  
 Gly Thr Thr Val Tyr Pro Asp Phe Phe Arg Glu Asp Thr Arg Glu Trp  
 305                            310                            315                            320  
 Trp Ala Gly Leu Ile Ser Glu Trp Leu Ser Gln Gly Val Asp Gly Ile  
                             325                            330                            335  
 Trp Leu Asp Met Asn Glu Pro Thr Asp Phe Ser Arg Ala Ile Glu Ile  
                             340                            345                            350  
 Arg Asp Val Leu Ser Ser Leu Pro Val Gln Phe Arg Asp Asp Arg Leu  
                             355                            360                            365  
 Val Thr Thr Phe Pro Asp Asn Val Val His Tyr Leu Arg Gly Lys Arg  
                             370                            375                            380  
 Val Lys His Glu Lys Val Arg Asn Ala Tyr Pro Leu Tyr Glu Ala Met  
 385                            390                            395                            400  
 Ala Thr Phe Lys Gly Phe Arg Thr Ser His Arg Asn Glu Ile Phe Ile  
                             405                            410                            415  
 Leu Ser Arg Ala Gly Tyr Ala Gly Ile Gln Arg Tyr Ala Phe Ile Trp  
                             420                            425                            430  
 Thr Gly Asp Asn Thr Pro Ser Trp Asp Asp Leu Lys Leu Gln Leu Gln  
                             435                            440                            445  
 Leu Val Leu Gly Leu Ser Ile Ser Gly Val Pro Phe Val Gly Cys Asp  
                             450                            455                            460  
 Ile Gly Gly Phe Gln Gly Arg Asn Phe Ala Glu Ile Asp Asn Ser Met  
 465                            470                            475                            480  
 Asp Leu Leu Val Lys Tyr Tyr Ala Leu Ala Leu Phe Phe Pro Phe Tyr  
                             485                            490                            495  
 Arg Ser His Lys Ala Thr Asp Gly Ile Asp Thr Glu Pro Val Phe Leu  
                             500                            505                            510

Pro Asp Tyr Tyr Lys Glu Lys Val Lys Glu Ile Val Glu Leu Arg Tyr  
 515 520 525  
 Lys Phe Leu Pro Tyr Ile Tyr Ser Leu Ala Leu Glu Ala Ser Glu Lys  
 530 535 540  
 Gly His Pro Val Ile Arg Pro Leu Phe Tyr Glu Phe Gln Asp Asp Asp  
 545 550 555 560  
 Asp Met Tyr Arg Ile Glu Asp Glu Tyr Met Val Gly Lys Tyr Leu Leu  
 565 570 575  
 Tyr Ala Pro Ile Val Ser Lys Glu Glu Ser Arg Leu Val Thr Leu Pro  
 580 585 590  
 Arg Gly Lys Trp Tyr Asn Tyr Trp Asn Gly Glu Ile Ile Asn Gly Lys  
 595 600 605  
 Ser Val Val Lys Ser Thr His Glu Leu Pro Ile Tyr Leu Arg Glu Gly  
 610 615 620  
 Ser Ile Ile Pro Leu Glu Gly Asp Glu Leu Ile Val Tyr Gly Glu Thr  
 625 630 635 640  
 Ser Phe Lys Arg Tyr Asp Asn Ala Glu Ile Thr Ser Ser Ser Asn Glu  
 645 650 655  
 Ile Lys Phe Ser Arg Glu Ile Tyr Val Ser Lys Leu Thr Ile Thr Ser  
 660 665 670  
 Glu Lys Pro Val Ser Lys Ile Ile Val Asp Asp Ser Lys Glu Ile Gln  
 675 680 685  
 Val Glu Lys Thr Met Gln Asn Thr Tyr Val Ala Lys Ile Asn Gln Lys  
 690 695 700  
 Ile Arg Gly Lys Ile Asn Leu Glu Ser Glu Lys Asp Glu Leu  
 705 710 715

&lt;210&gt; 37

&lt;211&gt; 1434

&lt;212&gt; DNA

&lt;213&gt; Thermotoga maritima

&lt;400&gt; 37

atgaagaaaa ccgctgctgc taaattcgaa cgccagcaca tggacagccc agatctgggt 60  
 accctggtgc cagcggttc catggccgag ttcttcccg agatcccga gatccagttc 120  
 gagggcaagg agtccaccaa cccgctcgcc ttccgcttct acgacccga cgaggtgac 180  
 gacggcaagc cgctcaagga ccacctcaag ttctccgtgg ccttctggca cacttcgtg 240  
 aacgagggcc gcgaccggt cggcgaccgc accgcccagc gcccgaggaa ccgcttctcc 300  
 gacccgatgg acaaggcctt cgcccgcgtg gacgcccctt tcgagttctg cgagaagctc 360  
 aacatcgagt acttctgctt ccacgaccgc gacatcgccc cggagggcaa gaccctccgc 420  
 gagaccaaca agatcctcga caaggtggtg gagcgcatca aggagcgcat gaaggactcc 480  
 aacgtgaagc tcctctgggg caccgccaac ctcttctccc acccgcgcta catgcacggc 540  
 gccgccacca cctgctccgc cgacgtgttc gcctacgccc ccgcccaggt gaagaaggcc 600  
 ctggagatca ccaaggagct gggcggcgag ggctacgtgt tctggggcgg ccgagaggc 660  
 tacgagaccc tcctcaacac cgacctcggc ctggagctgg agaacctcgc ccgcttcttc 720  
 cgcatggccg tggagtacgc caagaagatc ggcttcaccg gccagttcct catcgagccg 780  
 aagccgaagg agccgaccaa gcaccagtac gacttcgacg tggccaccgc ctacgccttc 840  
 ctcaagaacc acggcctcga cgagtacttc aagttcaaca tcgaggccaa ccacgccacc 900  
 ctgcgcggcc acaccttcca gcacgagctg cgcatggccc gcatectcgg caagctcggc 960  
 tccatcgacg ccaaccaggg cgacctcctc ctcggtggg acaccgacca gttcccagc 1020  
 aacatctacg acaccacct cgccatgtac gaggtgatca aggccggcg cttaccaag 1080  
 ggcggcctca acttcgacgc caaggtgcgc cgccctcctt acaaggtgga ggacctcttc 1140  
 atcggccaca tcgcccgcgt ggacaccttc gccctcggt tcaagatcgc ctacaagctc 1200

```

gccaaaggacg gcgtgttcga caagttcatc gaggagaagt accgctcctt caaggagggc 1260
atcggaagg agatcgtgga gggcaagacc gacttcgaga agctggagga gtacatcatc 1320
gacaaggagg acatcgagct gccgtccggc aagcaggagt acctggagtc cctcctcaac 1380
tcctacatcg tgaagaccat cgccgagctg cgctccgaga aggacgagct gtga 1434

```

&lt;210&gt; 38

&lt;211&gt; 477

&lt;212&gt; PRT

&lt;213&gt; Thermotoga maritima

&lt;400&gt; 38

```

Met Lys Glu Thr Ala Ala Ala Lys Phe Glu Arg Gln His Met Asp Ser
 1          5          10          15
Pro Asp Leu Gly Thr Leu Val Pro Arg Gly Ser Met Ala Glu Phe Phe
 20          25          30
Pro Glu Ile Pro Lys Ile Gln Phe Glu Gly Lys Glu Ser Thr Asn Pro
 35          40          45
Leu Ala Phe Arg Phe Tyr Asp Pro Asn Glu Val Ile Asp Gly Lys Pro
 50          55          60
Leu Lys Asp His Leu Lys Phe Ser Val Ala Phe Trp His Thr Phe Val
 65          70          75          80
Asn Glu Gly Arg Asp Pro Phe Gly Asp Pro Thr Ala Glu Arg Pro Trp
 85          90          95
Asn Arg Phe Ser Asp Pro Met Asp Lys Ala Phe Ala Arg Val Asp Ala
100          105          110
Leu Phe Glu Phe Cys Glu Lys Leu Asn Ile Glu Tyr Phe Cys Phe His
115          120          125
Asp Arg Asp Ile Ala Pro Glu Gly Lys Thr Leu Arg Glu Thr Asn Lys
130          135          140
Ile Leu Asp Lys Val Val Glu Arg Ile Lys Glu Arg Met Lys Asp Ser
145          150          155          160
Asn Val Lys Leu Leu Trp Gly Thr Ala Asn Leu Phe Ser His Pro Arg
165          170          175
Tyr Met His Gly Ala Ala Thr Thr Cys Ser Ala Asp Val Phe Ala Tyr
180          185          190
Ala Ala Ala Gln Val Lys Lys Ala Leu Glu Ile Thr Lys Glu Leu Gly
195          200          205
Gly Glu Gly Tyr Val Phe Trp Gly Gly Arg Glu Gly Tyr Glu Thr Leu
210          215          220
Leu Asn Thr Asp Leu Gly Leu Glu Leu Glu Asn Leu Ala Arg Phe Leu
225          230          235          240
Arg Met Ala Val Glu Tyr Ala Lys Lys Ile Gly Phe Thr Gly Gln Phe
245          250          255
Leu Ile Glu Pro Lys Pro Lys Glu Pro Thr Lys His Gln Tyr Asp Phe
260          265          270
Asp Val Ala Thr Ala Tyr Ala Phe Leu Lys Asn His Gly Leu Asp Glu
275          280          285
Tyr Phe Lys Phe Asn Ile Glu Ala Asn His Ala Thr Leu Ala Gly His
290          295          300
Thr Phe Gln His Glu Leu Arg Met Ala Arg Ile Leu Gly Lys Leu Gly
305          310          315          320
Ser Ile Asp Ala Asn Gln Gly Asp Leu Leu Glu Gly Trp Asp Thr Asp
325          330          335
Gln Phe Pro Thr Asn Ile Tyr Asp Thr Thr Leu Ala Met Tyr Glu Val

```

	340		345		350										
Ile	Lys	Ala	Gly	Gly	Phe	Thr	Lys	Gly	Gly	Leu	Asn	Phe	Asp	Ala	Lys
	355						360					365			
Val	Arg	Arg	Ala	Ser	Tyr	Lys	Val	Glu	Asp	Leu	Phe	Ile	Gly	His	Ile
	370					375						380			
Ala	Gly	Met	Asp	Thr	Phe	Ala	Leu	Gly	Phe	Lys	Ile	Ala	Tyr	Lys	Leu
385					390					395				400	
Ala	Lys	Asp	Gly	Val	Phe	Asp	Lys	Phe	Ile	Glu	Glu	Lys	Tyr	Arg	Ser
			405						410				415		
Phe	Lys	Glu	Gly	Ile	Gly	Lys	Glu	Ile	Val	Glu	Gly	Lys	Thr	Asp	Phe
			420					425					430		
Glu	Lys	Leu	Glu	Glu	Tyr	Ile	Ile	Asp	Lys	Glu	Asp	Ile	Glu	Leu	Pro
	435						440					445			
Ser	Gly	Lys	Gln	Glu	Tyr	Leu	Glu	Ser	Leu	Leu	Asn	Ser	Tyr	Ile	Val
	450					455					460				
Lys	Thr	Ile	Ala	Glu	Leu	Arg	Ser	Glu	Lys	Asp	Glu	Leu			
465					470					475					

&lt;210&gt; 39

&lt;211&gt; 1434

&lt;212&gt; DNA

&lt;213&gt; Thermotoga neapolitana

&lt;400&gt; 39

```

atgaaagaaa  ccgctgctgc  taaattcgaa  cgccagcaca  tggacagccc  agatctgggt  60
accctgggtgc  cagcggggttc  catggccgag  ttcttcccgg  agatcccgaa  ggtgcagttc  120
gagggcaagg  agtccaccaa  cccgctcgcc  ttcaagttct  acgacccgga  ggagatcatc  180
gacggcaagc  cgctcaagga  ccacctcaag  ttctccgtgg  ccttctggca  caccttcgtg  240
aacgagggcc  gcgaccggtt  cgcgaccgg  accgcccacc  gcccgaggaa  ccgctacacc  300
gacccgatgg  acaaggcctt  cgcccgcgtg  gacgcccctt  tcgagttctg  cgagaagctc  360
aacatcgagt  acttctgctt  ccacgaccgc  gacatcgccc  cggagggcaa  gaccctccgc  420
gagaccaaca  agatcctcga  caaggtggtg  gagcgcatca  aggagcgcat  gaaggactcc  480
aacgtgaagc  tcctctgggg  caccgccaac  ctcttctccc  acccgcgcta  catgcacggc  540
gccgccacca  cctgctccgc  cgacgtgttc  gcctacgccc  ccgcccaggt  gaagaaggcc  600
ctggagatca  ccaaggagct  gggcggcgag  ggctacgtgt  tctggggcgg  ccgcgagggc  660
tacgagacc  tcctcaacac  cgacctcggc  ttcgagctgg  agaacctcgc  ccgcttcctc  720
cgcatggccg  tggactacgc  caagcgcatc  ggcttcaccg  gccagttcct  catcgagccg  780
aagccgaagg  agccgaccaa  gcaccagtac  gacttcgacg  tggccaccgc  ctacgccttc  840
ctcaagtc  acggcctcga  cgagtacttc  aagttcaaca  tcgaggccaa  ccacgccacc  900
ctcgccggcc  acaccttcca  gcacgagctg  cgcatggccc  gcatcctcgg  caagctcggc  960
tccatcgacg  ccaaccaggg  cgacctcttc  ctgggctggg  acaccgacca  gttcccggac  1020
aacgtgtacg  acaccaccct  cgccatgtac  gaggtgatca  aggccggcgg  cttaccaag  1080
ggcgccctca  acttcgacgc  caaggtgcgc  cgcgccctct  acaaggtgga  ggacctcttc  1140
atcgccaca  tcgcccgcgt  ggacaccttc  gccctcggtc  tcaaggtggc  ctacaagctc  1200
gtgaaggacg  gcgtgctcga  caagttcatc  gaggagaagt  accgctcctt  ccgcgagggc  1260
atcgccgcg  acatcggtga  gggcaagggt  gacttcgaga  agctggagga  gtacatcatc  1320
gacaaggaga  ccacgcagct  gccgtccggc  aagcaggagt  acctggagtc  cctcatcaac  1380
tcctacatcg  tgaagacat  cctggagctg  cgctccgaga  aggacgagct  gtga  1434

```

&lt;210&gt; 40

&lt;211&gt; 477

&lt;212&gt; PRT

&lt;213&gt; Thermotoga neapolitana

&lt;400&gt; 40

```

Met Lys Glu Thr Ala Ala Lys Phe Glu Arg Gln His Met Asp Ser
 1          5          10          15
Pro Asp Leu Gly Thr Leu Val Pro Arg Gly Ser Met Ala Glu Phe Phe
      20          25          30
Pro Glu Ile Pro Lys Val Gln Phe Glu Gly Lys Glu Ser Thr Asn Pro
      35          40          45
Leu Ala Phe Lys Phe Tyr Asp Pro Glu Glu Ile Ile Asp Gly Lys Pro
 50          55          60
Leu Lys Asp His Leu Lys Phe Ser Val Ala Phe Trp His Thr Phe Val
 65          70          75          80
Asn Glu Gly Arg Asp Pro Phe Gly Asp Pro Thr Ala Asp Arg Pro Trp
      85          90          95
Asn Arg Tyr Thr Asp Pro Met Asp Lys Ala Phe Ala Arg Val Asp Ala
      100          105          110
Leu Phe Glu Phe Cys Glu Lys Leu Asn Ile Glu Tyr Phe Cys Phe His
      115          120          125
Asp Arg Asp Ile Ala Pro Glu Gly Lys Thr Leu Arg Glu Thr Asn Lys
      130          135          140
Ile Leu Asp Lys Val Val Glu Arg Ile Lys Glu Arg Met Lys Asp Ser
 145          150          155          160
Asn Val Lys Leu Leu Trp Gly Thr Ala Asn Leu Phe Ser His Pro Arg
      165          170          175
Tyr Met His Gly Ala Ala Thr Thr Cys Ser Ala Asp Val Phe Ala Tyr
      180          185          190
Ala Ala Ala Gln Val Lys Lys Ala Leu Glu Ile Thr Lys Glu Leu Gly
      195          200          205
Gly Glu Gly Tyr Val Phe Trp Gly Gly Arg Glu Gly Tyr Glu Thr Leu
      210          215          220
Leu Asn Thr Asp Leu Gly Phe Glu Leu Glu Asn Leu Ala Arg Phe Leu
 225          230          235          240
Arg Met Ala Val Asp Tyr Ala Lys Arg Ile Gly Phe Thr Gly Gln Phe
      245          250          255
Leu Ile Glu Pro Lys Pro Lys Glu Pro Thr Lys His Gln Tyr Asp Phe
      260          265          270
Asp Val Ala Thr Ala Tyr Ala Phe Leu Lys Ser His Gly Leu Asp Glu
      275          280          285
Tyr Phe Lys Phe Asn Ile Glu Ala Asn His Ala Thr Leu Ala Gly His
      290          295          300
Thr Phe Gln His Glu Leu Arg Met Ala Arg Ile Leu Gly Lys Leu Gly
 305          310          315          320
Ser Ile Asp Ala Asn Gln Gly Asp Leu Leu Leu Gly Trp Asp Thr Asp
      325          330          335
Gln Phe Pro Thr Asn Val Tyr Asp Thr Thr Leu Ala Met Tyr Glu Val
      340          345          350
Ile Lys Ala Gly Gly Phe Thr Lys Gly Gly Leu Asn Phe Asp Ala Lys
      355          360          365
Val Arg Arg Ala Ser Tyr Lys Val Glu Asp Leu Phe Ile Gly His Ile
      370          375          380
Ala Gly Met Asp Thr Phe Ala Leu Gly Phe Lys Val Ala Tyr Lys Leu
 385          390          395          400
Val Lys Asp Gly Val Leu Asp Lys Phe Ile Glu Glu Lys Tyr Arg Ser
      405          410          415

```



Phe Arg Glu Gly Ile Gly Arg Asp Ile Val Glu Gly Lys Val Asp Phe  
                   420                  425                  430  
 Glu Lys Leu Glu Glu Tyr Ile Ile Asp Lys Glu Thr Ile Glu Leu Pro  
                   435                  440                  445  
 Ser Gly Lys Gln Glu Tyr Leu Glu Ser Leu Ile Asn Ser Tyr Ile Val  
                   450                  455                  460  
 Lys Thr Ile Leu Glu Leu Arg Ser Glu Lys Asp Glu Leu  
 465                  470                  475

<210> 41  
 <211> 1435  
 <212> DNA  
 <213> *Thermotoga maritima*

<400> 41  
 atgggcagca gccatcatca tcatcatcac agcagcggcc tgggtgccgcg cggcagccat 60  
 atggctagca tgactgggtg acagcaaatg ggtcggatcc ccatggccga gttcttcccg 120  
 gagatcccga agatccagtt cgaggggcaag gagtccacca acccgctcgc ctcccgcttc 180  
 tacgacccga acgaggtgat cgacggcaag ccgctcaagg accacctcaa gttctccgtg 240  
 gccttctggc acaccttcgt gaacgagggc cgcgaccctg tcggcgaccc gaccgcccag 300  
 cgcccgtaga acgcgttctc cgacccgatg gacaaggcct tcgcccgcgt ggacgccctc 360  
 ttcgagttct gcgagaagct caacatcgag tacttctgct tccacgaccg cgacatcccc 420  
 cggaggggcaa gaccctccgc gagaccaaca agatcctcga caagggtggg gagcgcatca 480  
 aggagcgcac gaaggactcc aacgtgaagc tcctctgggg caccgccaac ctcttctccc 540  
 acccgcgcta catgcacggc gccgccacca cctgctccgc cgacgtgttc gcctacgccg 600  
 ccgcccagggt gaagaaggcc ctggagatca ccaaggagct gggcggcgag ggctacgtgt 660  
 tctggggcgg ccgcgagggc tacgagaccc tcctcaacac cgacctcggc ctggagctgg 720  
 agaacctcgc ccgcttcttc cgcagtgccg tggagtacgc caagaagatc ggcttcaccg 780  
 gccagttcct catcgagccg aagccgaagg agccgaccaa gcaccagtag gcttcgacgt 840  
 ggccaccgcc tacgccttcc tcaagaacca cggcctcgac gagtacttca agttcaacat 900  
 cgaggccaac cagccacccc tcgcccggca cacttccag cagcagctgc gcatggcccc 960  
 catcctcggc aagctcggct ccacgcgc caaccagggc gacctcctcc tcggctggga 1020  
 caccgaccag ttcccgaacca acatctacga caccacctc gccatgtacg aggtgatcaa 1080  
 ggccggcggc ttccaccaagg gcggcctcaa cttegcgc aaggtgcgcc gcgcctccta 1140  
 caagggtggag gacctcttca tcggccacat cgcggcatg gacaccttcg cctcggctt 1200  
 caagatcgcc tacaagctcg ccaaggacgg cgtgttcgac aagttcatcg aggagaagta 1260  
 ccgctccttc aaggagggca tcggcaagga gatcgtggag ggcaagaccg acttcgagaa 1320  
 gctggaggag tacatcatcg acaaggagga catcgagctg ccgtccggca agcaggagta 1380  
 cctggagctc ctctcaact cctacatcgt gaagaccatc gccgagctgc gctga 1435

<210> 42  
 <211> 478  
 <212> PRT  
 <213> *Thermotoga maritima*

<400> 42  
 Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro  
   1                  5                  10                  15  
 Arg Gly Ser His Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg  
                   20                  25                  30  
 Ile Pro Met Ala Glu Phe Phe Pro Glu Ile Pro Lys Ile Gln Phe Glu  
                   35                  40                  45  
 Gly Lys Glu Ser Thr Asn Pro Leu Ala Phe Arg Phe Tyr Asp Pro Asn

50		55		60
Glu Val Ile Asp Gly Lys Pro Leu Lys Asp His Leu Lys Phe Ser Val				
65		70		75
Ala Phe Trp His Thr Phe Val Asn Glu Gly Arg Asp Pro Phe Gly Asp				80
	85		90	95
Pro Thr Ala Glu Arg Pro Trp Asn Arg Phe Ser Asp Pro Met Asp Lys				
	100		105	110
Ala Phe Ala Arg Val Asp Ala Leu Phe Glu Phe Cys Glu Lys Leu Asn				
	115		120	125
Ile Glu Tyr Phe Cys Phe His Asp Arg Asp Ile Ala Pro Glu Gly Lys				
	130		135	140
Thr Leu Arg Glu Thr Asn Lys Ile Leu Asp Lys Val Val Glu Arg Ile				
	145		150	155
Lys Glu Arg Met Lys Asp Ser Asn Val Lys Leu Leu Trp Gly Thr Ala				
	165		170	175
Asn Leu Phe Ser His Pro Arg Tyr Met His Gly Ala Ala Thr Thr Cys				
	180		185	190
Ser Ala Asp Val Phe Ala Tyr Ala Ala Ala Gln Val Lys Lys Ala Leu				
	195		200	205
Glu Ile Thr Lys Glu Leu Gly Gly Glu Gly Tyr Val Phe Trp Gly Gly				
	210		215	220
Arg Glu Gly Tyr Glu Thr Leu Leu Asn Thr Asp Leu Gly Leu Glu Leu				
	225		230	235
Glu Asn Leu Ala Arg Phe Leu Arg Met Ala Val Glu Tyr Ala Lys Lys				
	245		250	255
Ile Gly Phe Thr Gly Gln Phe Leu Ile Glu Pro Lys Pro Lys Glu Pro				
	260		265	270
Thr Lys His Gln Tyr Asp Phe Asp Val Ala Thr Ala Tyr Ala Phe Leu				
	275		280	285
Lys Asn His Gly Leu Asp Glu Tyr Phe Lys Phe Asn Ile Glu Ala Asn				
	290		295	300
His Ala Thr Leu Ala Gly His Thr Phe Gln His Glu Leu Arg Met Ala				
	305		310	315
Arg Ile Leu Gly Lys Leu Gly Ser Ile Asp Ala Asn Gln Gly Asp Leu				
	325		330	335
Leu Leu Gly Trp Asp Thr Asp Gln Phe Pro Thr Asn Ile Tyr Asp Thr				
	340		345	350
Thr Leu Ala Met Tyr Glu Val Ile Lys Ala Gly Gly Phe Thr Lys Gly				
	355		360	365
Gly Leu Asn Phe Asp Ala Lys Val Arg Arg Ala Ser Tyr Lys Val Glu				
	370		375	380
Asp Leu Phe Ile Gly His Ile Ala Gly Met Asp Thr Phe Ala Leu Gly				
	385		390	395
Phe Lys Ile Ala Tyr Lys Leu Ala Lys Asp Gly Val Phe Asp Lys Phe				
	405		410	415
Ile Glu Glu Lys Tyr Arg Ser Phe Lys Glu Gly Ile Gly Lys Glu Ile				
	420		425	430
Val Glu Gly Lys Thr Asp Phe Glu Lys Leu Glu Glu Tyr Ile Ile Asp				
	435		440	445
Lys Glu Asp Ile Glu Leu Pro Ser Gly Lys Gln Glu Tyr Leu Glu Ser				
	450		455	460
Leu Leu Asn Ser Tyr Ile Val Lys Thr Ile Ala Glu Leu Arg				
	465		470	475

<210> 43  
 <211> 1436  
 <212> DNA  
 <213> *Thermotoga neapolitana*

<400> 43  
 atgggcagca gccatcatca tcatcatcac agcagcggcc tgggtgccgcg cggcagccat 60  
 atggctagca tgactggtgg acagcaaagtg ggtcggatcc ccatggccga gttcttcccc 120  
 gagatcccga aggtgcagtt cgagggcaag gattccacca acccgctcgc cttcaagtgc 180  
 tacgacccgg aggagatcat cgacggcaag ccgctcaagg accacctcaa gttctccgtg 240  
 gccttctggc acaccttcgt gaacgagggc cgcgaccgtg tcggcgaccc gaccgcccgc 300  
 cgcgcgtgga accgctacac cgaccgatg gacaaggcct tcgcccgcgt ggacgccttc 360  
 ttcgagttct gcgagaagct caacatcgag tacttctgct tccacgaccg cgacatcccc 420  
 cggagggcaa gaccttcgc gagaccaaca agatcctcga caaggtggtg gagcgcatca 480  
 aggagcgcat gaaggactcc aacgtgaagc tcctctgggg caccgccaac ctcttctccc 540  
 accgcgccta catgcacggc gccgccacca cctgctccgc cgacgtgttc gcctacgccg 600  
 ccgcccaggt gaagaaggcc ctggagatca ccaaggagct gggcggcgag ggctacgtgt 660  
 tctggggcgg ccgcgagggc tacgagaccc tcctcaacac cgacctcggc ttcgagctgg 720  
 agaacctcgc ccgcttcttc cgcattggccg tggactacgc caagcgcatc ggcttcaccg 780  
 gccagttcct catcgagccg aagcgaagg agccgaccaa gcaccagtac gacttcgacg 840  
 tggccaccgc ctacgccttc ctcaagtccc acggcctcga cgagtacttc aagttcaaca 900  
 tcgaggccaa ccacgccacc ctcgccggcc acaccttcca gcacgagctg cgcattggccc 960  
 gcatcctcgg caagctcggc tccatcgacg ccaaccaggg cgacctcttc ctcggtctgg 1020  
 acaccgacca gttcccgcacc aacgtgtacg acaccacct cgccatgtac gaggtgatca 1080  
 aggcggcgcg cttaccacaag ggcggcctca acttcgacgc caaggtgcgc cgcgcctcct 1140  
 acaagggtgga ggacctcttc atcgccaca tcgcccgcac ggacaccttc gccctcggct 1200  
 tcaagggtggc ctacaagctc gtgaaggacg gcgtgctcga caagttcatc gaggagaagt 1260  
 accgctcctt ccgcgagggc atcgccgcgc acatcggtga gggcaagggtg gacttcgaga 1320  
 agctggagga gtacatcatc gacaaggaga ccacgagctg gccgtccggc aagcaggagt 1380  
 acctggagtc cctcatcaac tcctacatcg tgaagacat cctggagctg cgctga 1436

<210> 44  
 <211> 478  
 <212> PRT  
 <213> *Thermotoga neapolitana*

<400> 44  
 Met Gly Ser Ser His His His His Ser Ser Gly Leu Val Pro  
 1 5 10 15  
 Arg Gly Ser His Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg  
 20 25 30  
 Ile Pro Met Ala Glu Phe Phe Pro Glu Ile Pro Lys Val Gln Phe Glu  
 35 40 45  
 Gly Lys Glu Ser Thr Asn Pro Leu Ala Phe Lys Phe Tyr Asp Pro Glu  
 50 55 60  
 Glu Ile Ile Asp Gly Lys Pro Leu Lys Asp His Leu Lys Phe Ser Val  
 65 70 75 80  
 Ala Phe Trp His Thr Phe Val Asn Glu Gly Arg Asp Pro Phe Gly Asp  
 85 90 95  
 Pro Thr Ala Asp Arg Pro Trp Asn Arg Tyr Thr Asp Pro Met Asp Lys  
 100 105 110  
 Ala Phe Ala Arg Val Asp Ala Leu Phe Glu Phe Cys Glu Lys Leu Asn  
 115 120 125

```

Ile Glu Tyr Phe Cys Phe His Asp Arg Asp Ile Ala Pro Glu Gly Lys
  130                      135                      140
Thr Leu Arg Glu Thr Asn Lys Ile Leu Asp Lys Val Val Glu Arg Ile
145                      150                      155                      160
Lys Glu Arg Met Lys Asp Ser Asn Val Lys Leu Leu Trp Gly Thr Ala
                      165                      170                      175
Asn Leu Phe Ser His Pro Arg Tyr Met His Gly Ala Ala Thr Thr Cys
                      180                      185                      190
Ser Ala Asp Val Phe Ala Tyr Ala Ala Ala Gln Val Lys Lys Ala Leu
                      195                      200                      205
Glu Ile Thr Lys Glu Leu Gly Gly Glu Gly Tyr Val Phe Trp Gly Gly
  210                      215                      220
Arg Glu Gly Tyr Glu Thr Leu Leu Asn Thr Asp Leu Gly Phe Glu Leu
225                      230                      235                      240
Glu Asn Leu Ala Arg Phe Leu Arg Met Ala Val Asp Tyr Ala Lys Arg
                      245                      250                      255
Ile Gly Phe Thr Gly Gln Phe Leu Ile Glu Pro Lys Pro Lys Glu Pro
                      260                      265                      270
Thr Lys His Gln Tyr Asp Phe Asp Val Ala Thr Ala Tyr Ala Phe Leu
                      275                      280                      285
Lys Ser His Gly Leu Asp Glu Tyr Phe Lys Phe Asn Ile Glu Ala Asn
  290                      295                      300
His Ala Thr Leu Ala Gly His Thr Phe Gln His Glu Leu Arg Met Ala
305                      310                      315                      320
Arg Ile Leu Gly Lys Leu Gly Ser Ile Asp Ala Asn Gln Gly Asp Leu
                      325                      330                      335
Leu Leu Gly Trp Asp Thr Asp Gln Phe Pro Thr Asn Val Tyr Asp Thr
  340                      345                      350
Thr Leu Ala Met Tyr Glu Val Ile Lys Ala Gly Gly Phe Thr Lys Gly
                      355                      360                      365
Gly Leu Asn Phe Asp Ala Lys Val Arg Arg Ala Ser Tyr Lys Val Glu
  370                      375                      380
Asp Leu Phe Ile Gly His Ile Ala Gly Met Asp Thr Phe Ala Leu Gly
385                      390                      395                      400
Phe Lys Val Ala Tyr Lys Leu Val Lys Asp Gly Val Leu Asp Lys Phe
                      405                      410                      415
Ile Glu Glu Lys Tyr Arg Ser Phe Arg Glu Gly Ile Gly Arg Asp Ile
  420                      425                      430
Val Glu Gly Lys Val Asp Phe Glu Lys Leu Glu Glu Tyr Ile Ile Asp
  435                      440                      445
Lys Glu Thr Ile Glu Leu Pro Ser Gly Lys Gln Glu Tyr Leu Glu Ser
  450                      455                      460
Leu Ile Asn Ser Tyr Ile Val Lys Thr Ile Leu Glu Leu Arg
465                      470                      475

```

&lt;210&gt; 45

&lt;211&gt; 1095

&lt;212&gt; PRT

<213> *Aspergillus shirousami*

&lt;400&gt; 45

```

Ala Thr Pro Ala Asp Trp Arg Ser Gln Ser Ile Tyr Phe Leu Leu Thr
  1                      5                      10                      15

```

Asp Arg Phe Ala Arg Thr Asp Gly Ser Thr Thr Ala Thr Cys Asn Thr  
 20 25 30  
 Ala Asp Gln Lys Tyr Cys Gly Gly Thr Trp Gln Gly Ile Ile Asp Lys  
 35 40 45  
 Leu Asp Tyr Ile Gln Gly Met Gly Phe Thr Ala Ile Trp Ile Thr Pro  
 50 55 60  
 Val Thr Ala Gln Leu Pro Gln Thr Thr Ala Tyr Gly Asp Ala Tyr His  
 65 70 75 80  
 Gly Tyr Trp Gln Gln Asp Ile Tyr Ser Leu Asn Glu Asn Tyr Gly Thr  
 85 90 95  
 Ala Asp Asp Leu Lys Ala Leu Ser Ser Ala Leu His Glu Arg Gly Met  
 100 105 110  
 Tyr Leu Met Val Asp Val Val Ala Asn His Met Gly Tyr Asp Gly Ala  
 115 120 125  
 Gly Ser Ser Val Asp Tyr Ser Val Phe Lys Pro Phe Ser Ser Gln Asp  
 130 135 140  
 Tyr Phe His Pro Phe Cys Phe Ile Gln Asn Tyr Glu Asp Gln Thr Gln  
 145 150 155 160  
 Val Glu Asp Cys Trp Leu Gly Asp Asn Thr Val Ser Leu Pro Asp Leu  
 165 170 175  
 Asp Thr Thr Lys Asp Val Val Lys Asn Glu Trp Tyr Asp Trp Val Gly  
 180 185 190  
 Ser Leu Val Ser Asn Tyr Ser Ile Asp Gly Leu Arg Ile Asp Thr Val  
 195 200 205  
 Lys His Val Gln Lys Asp Phe Trp Pro Gly Tyr Asn Lys Ala Ala Gly  
 210 215 220  
 Val Tyr Cys Ile Gly Glu Val Leu Asp Val Asp Pro Ala Tyr Thr Cys  
 225 230 235 240  
 Pro Tyr Gln Asn Val Met Asp Gly Val Leu Asn Tyr Pro Ile Tyr Tyr  
 245 250 255  
 Pro Leu Leu Asn Ala Phe Lys Ser Thr Ser Gly Ser Met Asp Asp Leu  
 260 265 270  
 Tyr Asn Met Ile Asn Thr Val Lys Ser Asp Cys Pro Asp Ser Thr Leu  
 275 280 285  
 Leu Gly Thr Phe Val Glu Asn His Asp Asn Pro Arg Phe Ala Ser Tyr  
 290 295 300  
 Thr Asn Asp Ile Ala Leu Ala Lys Asn Val Ala Ala Phe Ile Ile Leu  
 305 310 315 320  
 Asn Asp Gly Ile Pro Ile Ile Tyr Ala Gly Gln Glu Gln His Tyr Ala  
 325 330 335  
 Gly Gly Asn Asp Pro Ala Asn Arg Glu Ala Thr Trp Leu Ser Gly Tyr  
 340 345 350  
 Pro Thr Asp Ser Glu Leu Tyr Lys Leu Ile Ala Ser Ala Asn Ala Ile  
 355 360 365  
 Arg Asn Tyr Ala Ile Ser Lys Asp Thr Gly Phe Val Thr Tyr Lys Asn  
 370 375 380  
 Trp Pro Ile Tyr Lys Asp Asp Thr Thr Ile Ala Met Arg Lys Gly Thr  
 385 390 395 400  
 Asp Gly Ser Gln Ile Val Thr Ile Leu Ser Asn Lys Gly Ala Ser Gly  
 405 410 415  
 Asp Ser Tyr Thr Leu Ser Leu Ser Gly Ala Gly Tyr Thr Ala Gly Gln  
 420 425 430  
 Gln Leu Thr Glu Val Ile Gly Cys Thr Thr Val Thr Val Gly Ser Asp  
 435 440 445

Gly Asn Val Pro Val Pro Met Ala Gly Gly Leu Pro Arg Val Leu Tyr  
 450 455 460  
 Pro Thr Glu Lys Leu Ala Gly Ser Lys Ile Cys Ser Ser Ser Lys Pro  
 465 470 475 480  
 Ala Thr Leu Asp Ser Trp Leu Ser Asn Glu Ala Thr Val Ala Arg Thr  
 485 490 495  
 Ala Ile Leu Asn Asn Ile Gly Ala Asp Gly Ala Trp Val Ser Gly Ala  
 500 505 510  
 Asp Ser Gly Ile Val Val Ala Ser Pro Ser Thr Asp Asn Pro Asp Tyr  
 515 520 525  
 Phe Tyr Thr Trp Thr Arg Asp Ser Gly Ile Val Leu Lys Thr Leu Val  
 530 535 540  
 Asp Leu Phe Arg Asn Gly Asp Thr Asp Leu Leu Ser Thr Ile Glu His  
 545 550 555 560  
 Tyr Ile Ser Ser Gln Ala Ile Ile Gln Gly Val Ser Asn Pro Ser Gly  
 565 570 575  
 Asp Leu Ser Ser Gly Gly Leu Gly Glu Pro Lys Phe Asn Val Asp Glu  
 580 585 590  
 Thr Ala Tyr Ala Gly Ser Trp Gly Arg Pro Gln Arg Asp Gly Pro Ala  
 595 600 605  
 Leu Arg Ala Thr Ala Met Ile Gly Phe Gly Gln Trp Leu Leu Asp Asn  
 610 615 620  
 Gly Tyr Thr Ser Ala Ala Thr Glu Ile Val Trp Pro Leu Val Arg Asn  
 625 630 635 640  
 Asp Leu Ser Tyr Val Ala Gln Tyr Trp Asn Gln Thr Gly Tyr Asp Leu  
 645 650 655  
 Trp Glu Glu Val Asn Gly Ser Ser Phe Phe Thr Ile Ala Val Gln His  
 660 665 670  
 Arg Ala Leu Val Glu Gly Ser Ala Phe Ala Thr Ala Val Gly Ser Ser  
 675 680 685  
 Cys Ser Trp Cys Asp Ser Gln Ala Pro Gln Ile Leu Cys Tyr Leu Gln  
 690 695 700  
 Ser Phe Trp Thr Gly Ser Tyr Ile Leu Ala Asn Phe Asp Ser Ser Arg  
 705 710 715 720  
 Ser Gly Lys Asp Thr Asn Thr Leu Leu Gly Ser Ile His Thr Phe Asp  
 725 730 735  
 Pro Glu Ala Gly Cys Asp Asp Ser Thr Phe Gln Pro Cys Ser Pro Arg  
 740 745 750  
 Ala Leu Ala Asn His Lys Glu Val Val Asp Ser Phe Arg Ser Ile Tyr  
 755 760 765  
 Thr Leu Asn Asp Gly Leu Ser Asp Ser Glu Ala Val Ala Val Gly Arg  
 770 775 780  
 Tyr Pro Glu Asp Ser Tyr Tyr Asn Gly Asn Pro Trp Phe Leu Cys Thr  
 785 790 795 800  
 Leu Ala Ala Ala Glu Gln Leu Tyr Asp Ala Leu Tyr Gln Trp Asp Lys  
 805 810 815  
 Gln Gly Ser Leu Glu Ile Thr Asp Val Ser Leu Asp Phe Phe Lys Ala  
 820 825 830  
 Leu Tyr Ser Gly Ala Ala Thr Gly Thr Tyr Ser Ser Ser Ser Ser Thr  
 835 840 845  
 Tyr Ser Ser Ile Val Ser Ala Val Lys Thr Phe Ala Asp Gly Phe Val  
 850 855 860  
 Ser Ile Val Glu Thr His Ala Ala Ser Asn Gly Ser Leu Ser Glu Gln  
 865 870 875 880

<400>	46					
gccaccccg	cgactggcg	ctcccagtc	atctacttc	tcttcaccga	ccgcttcgcc	60
cgcaccgacg	gtccaccac	cgccacctgc	aacaccgcg	accagaagta	ctgcggcggc	120
acctggcagg	gcatcatcga	caagctcgac	tacatccagg	gcatgggctt	caccgcctac	180
tggatcaccc	cggtagccgc	ccagctcccg	cgaccaccg	cctacggcga	cgcttaccac	240
ggctactggc	agcaggacat	ctactccctc	aacgagaact	acggcacccg	cgcagccttc	300
aaggccctct	cctccgcctt	ccacgagcgc	ggcatgtacc	tcatggtgga	cgtggtggcc	360
aaccacatgg	gctacgacgg	cgccgggtcc	tccgtggact	actccgtgtt	caagccgttc	420
tcttcccagg	actacttcca	cccgttctgc	ttcatccaga	actacgagga	ccagacccag	480
gtggaggact	gctggtctcg	cgacaacacc	gtgtccctcc	cggacctcga	caccaccaag	540
gacgtggtga	agaacgagtg	gtacgactgc	gtgggtctcc	tctgttccaa	ctactccatc	600
gacggccttc	gcctcgacac	cgtgaagcac	gtgcagaagg	acttctggcc	gggtacaac	660
aaggccgcgc	gcgtgtactg	catcggcgag	gtgctcgacg	tggaccgggc	ctacacctgc	720
ccgtaccaga	acgtgatgga	cggcgtgtct	aactacccga	tctactacct	gctcctcaac	780
gccttcaagt	ccacctccgg	ctcgatggac	gacctctaca	acatgatcaa	caccgtgaag	840
tccgactgcc	cggactccac	cctcctcggc	accttcgtgg	agaaccacga	caaccgcgc	900
ttgcctctct	acaccaacga	ctatgccttc	gccaaagaac	tggccgcctt	catcatcttc	960
aacgcaggca	tcccagcatc	ctacgcggcg	caggagcagc	actacgcggg	cggcaacgac	1020
ccggccaacc	gcgaggccac	ctggctctcc	ggctaccgga	ccgactccga	cgtgtacaag	1080

```

ctcatcgect ccgccaacgc catccgcaac tacgccatct ccaaggacac cggcttcgtg 1140
acctacaaga actggccgat ctacaaggac gacaccacca tcgccaatgcg caagggcacc 1200
gacggctccc agatcgtgac catcctctcc aacaagggcg cctccggcga ctctacacc 1260
ctctccctct cggcgccgg ctagaccgcc ggccagcagc tcaccgaggt gatcggctgc 1320
accaccgtga ccgtgggctc cgacggcaac gtgccgggtgc cgatggccgg cggcctccc 1380
cgcgtgctct acccgaccga gaagctcgcc ggctccaaga tatgctcttc ctccaagccg 1440
gccaccctcg actcctggct ctccaacgag gccaccgtgg ccgcaccgc catcctcaac 1500
aacatcggcg ccgacggcgc ctgggtgtcc ggcccgact ccggcatcgt ggtggcctcc 1560
ccgtccaccg acaaccggga ctacttttac acctggaccc gcgactccgg catcgtgctc 1620
aagaccctcg tggacctctt ccgcaacggc gacaccgacc tctctccac catcgagcac 1680
tacatctcct ccagggcat catccagggc gtgtccaacc cgtccggcga cctctcctcc 1740
ggcggcctcg gcgagccgaa gttcaacgtg gacgagaccg cctacgccgg ctctggggc 1800
cgcccgagc gcgacggccc ggccctccgc gccaccgcca tgatcggctt cggccagtgg 1860
ctctcgaca acggctacac ctccgcccgc accgagatcg tgtggccgct cgtgcgcaac 1920
gacctctcct acgtggcca gtactggaac cagaccggct acgacctctg ggaggaggtg 1980
aacggctcct ccttcttcac catcgccgtg cagcaccgcg cctcgtgga gggtccgc 2040
ttcgccaccg ccgtggctc ctctgctcc tggcgact cccaggcccc gcagatcctc 2100
tgctacctcc agtcttctg gaccggctcc tacatctctg ccaacttcga ctctccgc 2160
tccggcaagg acaccaacac cctctcggc tccatccaca ccttcgacc ggaggccggc 2220
tgcgacgact ccaccttcca gccgtgtcc ccgcgcgcc tcgccaacca caaggaggtg 2280
gtggactcct tccgtccat ctacacctc aacgacggcc tctccgactc cgaggccgtg 2340
gcgtggggc gctaccggga ggactcctac tacaacggca acccgtggtt cctctgcacc 2400
ctgcgcgcg ccgagcagct ctacgacgcc ctctaccagt gggacaagca gggctccctg 2460
gagatcaccg acgtgtcct cgacttctt aaggccctct actccggcgc cgccaccggc 2520
acctactcct cctctcctc cactactcc tccatcgtgt ccgcctgaa gacctcgc 2580
gacggcttcg tgtccatcgt ggagaccac gccgcctcca acggtccct ctccgagcag 2640
ttcgacaagt ccgacggcga cgagctgtcc gcccgcgacc tcacctggtc ctacgccgcc 2700
ctctcaccg ccaacaaccg ccgcaactcc gtggtgcccgc cgtcctgggg cgagacctc 2760
gcctctccg tgcggggcac ctgcgcgcc acctccgct ccggcaccta ctctccgtg 2820
accgtgacct ctggcgctc catcgtggcc accggcgga ccaccaccac cgccaccacc 2880
accggctccg gcggcgtgac ctccacctc aagaccacca ccaccgctc caagacctc 2940
accaccacct cctccacctc ctgcaccacc ccgaccgccc tggccgtgac ctccgacctc 3000
accgccacca ccacctacgg cgagaacatc tacctcgtgg gctccatctc ccagctcggc 3060
gactgggaga cctccgacgg catcgccctc tccgcgaca agtacacctc ctccaaccgc 3120
ccgtggtagc tgaccgtgac cctcccgcc ggcgagtcct tcgagtacaa gttcatccgc 3180
gtggagtcgc acgactccgt ggagtgggag tccgaccga accgcgagta caccgtgccg 3240
caggcctgcg gcgagtcac cgccaccgtg accgacacct ggcgc 3285

```

&lt;210&gt; 47

&lt;211&gt; 679

&lt;212&gt; PRT

<213> *Thermoanaerobacterium thermosaccharolyticum*

&lt;400&gt; 47

```

Val Leu Ser Gly Cys Ser Asn Asn Val Ser Ser Ile Lys Ile Asp Arg
 1           5           10           15
Phe Asn Asn Ile Ser Ala Val Asn Gly Pro Gly Glu Glu Asp Thr Trp
      20           25           30
Ala Ser Ala Gln Lys Gln Gly Val Gly Thr Ala Asn Asn Tyr Val Ser
      35           40           45
Arg Val Trp Phe Thr Leu Ala Asn Gly Ala Ile Ser Glu Val Tyr Tyr
      50           55           60
Pro Thr Ile Asp Thr Ala Asp Val Lys Glu Ile Lys Phe Ile Val Thr
      65           70           75           80

```



Asp Gly Lys Ser Phe Val Ser Asp Glu Thr Lys Asp Ala Ile Ser Lys  
 85 90 95  
 Val Glu Lys Phe Thr Asp Lys Ser Leu Gly Tyr Lys Leu Val Asn Thr  
 100 105 110  
 Asp Lys Lys Gly Arg Tyr Arg Ile Thr Lys Glu Ile Phe Thr Asp Val  
 115 120 125  
 Lys Arg Asn Ser Leu Ile Met Lys Ala Lys Phe Glu Ala Leu Glu Gly  
 130 135 140  
 Ser Ile His Asp Tyr Lys Leu Tyr Leu Ala Tyr Asp Pro His Ile Lys  
 145 150 155 160  
 Asn Gln Gly Ser Tyr Asn Glu Gly Tyr Val Ile Lys Ala Asn Asn Asn  
 165 170 175  
 Glu Met Leu Met Ala Lys Arg Asp Asn Val Tyr Thr Ala Leu Ser Ser  
 180 185 190  
 Asn Ile Gly Trp Lys Gly Tyr Ser Ile Gly Tyr Tyr Lys Val Asn Asp  
 195 200 205  
 Ile Met Thr Asp Leu Asp Glu Asn Lys Gln Met Thr Lys His Tyr Asp  
 210 215 220  
 Ser Ala Arg Gly Asn Ile Ile Glu Gly Ala Glu Ile Asp Leu Thr Lys  
 225 230 235 240  
 Asn Ser Glu Phe Glu Ile Val Leu Ser Phe Gly Gly Ser Asp Ser Glu  
 245 250 255  
 Ala Ala Lys Thr Ala Leu Glu Thr Leu Gly Glu Asp Tyr Asn Asn Leu  
 260 265 270  
 Lys Asn Asn Tyr Ile Asp Glu Trp Thr Lys Tyr Cys Asn Thr Leu Asn  
 275 280 285  
 Asn Phe Asn Gly Lys Ala Asn Ser Leu Tyr Tyr Asn Ser Met Met Ile  
 290 295 300  
 Leu Lys Ala Ser Glu Asp Lys Thr Asn Lys Gly Ala Tyr Ile Ala Ser  
 305 310 315 320  
 Leu Ser Ile Pro Trp Gly Asp Gly Gln Arg Asp Asp Asn Thr Gly Gly  
 325 330 335  
 Tyr His Leu Val Trp Ser Arg Asp Leu Tyr His Val Ala Asn Ala Phe  
 340 345 350  
 Ile Ala Ala Gly Asp Val Asp Ser Ala Asn Arg Ser Leu Asp Tyr Leu  
 355 360 365  
 Ala Lys Val Val Lys Asp Asn Gly Met Ile Pro Gln Asn Thr Trp Ile  
 370 375 380  
 Ser Gly Lys Pro Tyr Trp Thr Ser Ile Gln Leu Asp Glu Gln Ala Asp  
 385 390 395 400  
 Pro Ile Ile Leu Ser Tyr Arg Leu Lys Arg Tyr Asp Leu Tyr Asp Ser  
 405 410 415  
 Leu Val Lys Pro Leu Ala Asp Phe Ile Ile Lys Ile Gly Pro Lys Thr  
 420 425 430  
 Gly Gln Glu Arg Trp Glu Glu Ile Gly Gly Tyr Ser Pro Ala Thr Met  
 435 440 445  
 Ala Ala Glu Val Ala Gly Leu Thr Cys Ala Ala Tyr Ile Ala Glu Gln  
 450 455 460  
 Asn Lys Asp Tyr Glu Ser Ala Gln Lys Tyr Gln Glu Lys Ala Asp Asn  
 465 470 475 480  
 Trp Gln Lys Leu Ile Asp Asn Leu Thr Tyr Thr Glu Asn Gly Pro Leu  
 485 490 495  
 Gly Asn Gly Gln Tyr Tyr Ile Arg Ile Ala Gly Leu Ser Asp Pro Asn  
 500 505 510

Ala Asp Phe Met Ile Asn Ile Ala Asn Gly Gly Gly Val Tyr Asp Gln  
           515                          520                          525  
 Lys Glu Ile Val Asp Pro Ser Phe Leu Glu Leu Val Arg Leu Gly Val  
           530                          535                          540  
 Lys Ser Ala Asp Asp Pro Lys Ile Leu Asn Thr Leu Lys Val Val Asp  
 545                          550                          555                          560  
 Ser Thr Ile Lys Val Asp Thr Pro Lys Gly Pro Ser Trp Tyr Arg Tyr  
                           565                          570                          575  
 Asn His Asp Gly Tyr Gly Glu Pro Ser Lys Thr Glu Leu Tyr His Gly  
                           580                          585                          590  
 Ala Gly Lys Gly Arg Leu Trp Pro Leu Leu Thr Gly Glu Arg Gly Met  
                           595                          600                          605  
 Tyr Glu Ile Ala Ala Gly Lys Asp Ala Thr Pro Tyr Val Lys Ala Met  
                           610                          615                          620  
 Glu Lys Phe Ala Asn Glu Gly Gly Ile Ile Ser Glu Gln Val Trp Glu  
 625                          630                          635                          640  
 Asp Thr Gly Leu Pro Thr Asp Ser Ala Ser Pro Leu Asn Trp Ala His  
                           645                          650                          655  
 Ala Glu Tyr Val Ile Leu Phe Ala Ser Asn Ile Glu His Lys Val Leu  
                           660                          665                          670  
 Asp Met Pro Asp Ile Val Tyr  
                           675

&lt;210&gt; 48

&lt;211&gt; 2037

&lt;212&gt; DNA

&lt;213&gt; Thermoanaerobacterium thermosaccharolyticum

&lt;220&gt;

&lt;223&gt; synthetic

&lt;400&gt; 48

gtgctctcgg gctgctccaa caacgtgtcc tccatcaaga tcgaccgctt caacaacatc 60  
 tccgccgtga acgcccggg cgaggaggac acctgggcct ccgccagaa gcagggcgtg 120  
 ggcaccgcca acaactacgt gtcccgcgtg tggttcaccc tcgccaacgg cgccatctcc 180  
 gaggtgtact acccgaccat cgacaccgcc gacgtgaagg agatcaagtt catcgtgacc 240  
 gacggcaagt ccttcgtgtc cgacgagacc aaggacgcca tctccaagggt ggagaagttc 300  
 accgacaagt ccttcggcta caagctcgtg aacaccgaca agaagggccg ctaccgcac 360  
 accaaggaaa tcttcaccga cgtgaagcgc aactccctca tcatgaaggc caagttcgag 420  
 gccctcgagg gctccatcca cgactacaag ctctacctcg cctacgaccc gcacatcaag 480  
 aaccagggct cctacaacga gggctacgtg atcaaggcca acaacaacga gatgctcatg 540  
 gccaaagcgc acaacgtgta caccgccctc tcctccaaca tcggctggaa gggctactcc 600  
 atcggctact acaaggtgaa cgacatcatg accgacctcg acgagaacaa gcagatgacc 660  
 aagcactacg actccgcccg cggaacatc atcgagggcg ccgagatcga cctcaccaag 720  
 aactccgagt tcgagatcgt gctctccttc ggcggtccg actccgaggc cgccaagacc 780  
 gccctcgaga ccttcggcga ggactacaac aacctcaaga acaactacat cgacgagtgg 840  
 accaagtact gcaacaccct caacaacttc aacggcaagg ccaactccct ctactacaac 900  
 tccatgatga tcctcaaggc ctccgaggac aagaccaaca agggcgcta catcgctcc 960  
 ctctccatcc cgtggggcga cggccagcgc gacgacaaca ccggcgcta ccacctcgtg 1020  
 tggctcccgcg acctctacca cgtggccaac gccttcacg ccgcccgcga cgtggactcc 1080  
 gccaacgcgt ccttcgacta cctcgccaag gtggtgaagg acaacggcat gatcccgag 1140  
 aacacctgga tctccggcaa accgtactgg acctccatcc agctcgacga gcaggccgac 1200  
 ccgatcatcc tctcctaccg cctcaagcgc tacgacctct acgactccct cgtgaagccg 1260

```

ctcgccgact tcacatcaa gatcgggccg aagaccggcc aggagcgctg ggaggagatc 1320
ggcggtact ccccgccac gatggccgc gaggtggccg gcctcacctg cgccgcctac 1380
atcgccgagc agaacaagga ctacgagtc gcccagaagt accaggagaa ggccgacaac 1440
tggcagaagc tcacgacaa cctcacctac accgagaacg gcccgctcgg caacggccag 1500
tactacatcc gcatcgccg cctctccgac ccgaacgccg acttcatgat caacatcgcc 1560
aacggcgccg gcgtgtacga ccagaaggag atcgtggacc cgtccttctt cgagctggtg 1620
cgctcgccg tgaagtccgc cgacgaccgc aagatcctca acacctcaa ggtggtggac 1680
tccaccatca aggtggacac cccgaagggc ccgtcctggt atcgctacaa ccacgacggc 1740
tacggcgagc cgtccaagac cgagctgtac cacggcgccg gcaagggccg cctctggccg 1800
ctcctcaccg gcgagcgccg catgtacgag atcgccgccc gcaaggacgc caccctgtac 1860
gtgaaggcga tggagaagtt cgccaacgag ggccgcatca tctccgagca ggtgtgggag 1920
gacaccggcc tcccgaaccg ctccgcctcc ccgtcaact gggccacgc cgagtacgtg 1980
atcctcttcg cctccaacat cgagcacaag gtgctcgaca tgccggacat cgtgtac 2037

```

&lt;210&gt; 49

&lt;211&gt; 579

&lt;212&gt; PRT

<213> *Rhizopus oryzae*

&lt;400&gt; 49

```

Ala Ser Ile Pro Ser Ser Ala Ser Val Gln Leu Asp Ser Tyr Asn Tyr
1          5          10          15
Asp Gly Ser Thr Phe Ser Gly Lys Ile Tyr Val Lys Asn Ile Ala Tyr
20          25          30
Ser Lys Lys Val Thr Val Ile Tyr Ala Asp Gly Ser Asp Asn Trp Asn
35          40          45
Asn Asn Gly Asn Thr Ile Ala Ala Ser Tyr Ser Ala Pro Ile Ser Gly
50          55          60
Ser Asn Tyr Glu Tyr Trp Thr Phe Ser Ala Ser Ile Asn Gly Ile Lys
65          70          75          80
Glu Phe Tyr Ile Lys Tyr Glu Val Ser Gly Lys Thr Tyr Tyr Asp Asn
85          90          95
Asn Asn Ser Ala Asn Tyr Gln Val Ser Thr Ser Lys Pro Thr Thr Thr
100          105          110
Thr Ala Thr Ala Thr Thr Thr Thr Ala Pro Ser Thr Ser Thr Thr Thr
115          120          125
Pro Pro Ser Arg Ser Glu Pro Ala Thr Phe Pro Thr Gly Asn Ser Thr
130          135          140
Ile Ser Ser Trp Ile Lys Lys Gln Glu Gly Ile Ser Arg Phe Ala Met
145          150          155          160
Leu Arg Asn Ile Asn Pro Pro Gly Ser Ala Thr Gly Phe Ile Ala Ala
165          170          175
Ser Leu Ser Thr Ala Gly Pro Asp Tyr Tyr Tyr Ala Trp Thr Arg Asp
180          185          190
Ala Ala Leu Thr Ser Asn Val Ile Val Tyr Glu Tyr Asn Thr Thr Leu
195          200          205
Ser Gly Asn Lys Thr Ile Leu Asn Val Leu Lys Asp Tyr Val Thr Phe
210          215          220
Ser Val Lys Thr Gln Ser Thr Ser Thr Val Cys Asn Cys Leu Gly Glu
225          230          235          240
Pro Lys Phe Asn Pro Asp Ala Ser Gly Tyr Thr Gly Ala Trp Gly Arg
245          250          255
Pro Gln Asn Asp Gly Pro Ala Glu Arg Ala Thr Thr Phe Ile Leu Phe
260          265          270

```

Ala Asp Ser Tyr Leu Thr Gln Thr Lys Asp Ala Ser Tyr Val Thr Gly  
 275 280 285  
 Thr Leu Lys Pro Ala Ile Phe Lys Asp Leu Asp Tyr Val Val Asn Val  
 290 295 300  
 Trp Ser Asn Gly Cys Phe Asp Leu Trp Glu Glu Val Asn Gly Val His  
 305 310 315 320  
 Phe Tyr Thr Leu Met Val Met Arg Lys Gly Leu Leu Leu Gly Ala Asp  
 325 330 335  
 Phe Ala Lys Arg Asn Gly Asp Ser Thr Arg Ala Ser Thr Tyr Ser Ser  
 340 345 350  
 Thr Ala Ser Thr Ile Ala Asn Lys Ile Ser Ser Phe Trp Val Ser Ser  
 355 360 365  
 Asn Asn Trp Ile Gln Val Ser Gln Ser Val Thr Gly Gly Val Ser Lys  
 370 375 380  
 Lys Gly Leu Asp Val Ser Thr Leu Leu Ala Ala Asn Leu Gly Ser Val  
 385 390 395 400  
 Asp Asp Gly Phe Phe Thr Pro Gly Ser Glu Lys Ile Leu Ala Thr Ala  
 405 410 415  
 Val Ala Val Glu Asp Ser Phe Ala Ser Leu Tyr Pro Ile Asn Lys Asn  
 420 425 430  
 Leu Pro Ser Tyr Leu Gly Asn Ser Ile Gly Arg Tyr Pro Glu Asp Thr  
 435 440 445  
 Tyr Asn Gly Asn Gly Asn Ser Gln Gly Asn Ser Trp Phe Leu Ala Val  
 450 455 460  
 Thr Gly Tyr Ala Glu Leu Tyr Tyr Arg Ala Ile Lys Glu Trp Ile Gly  
 465 470 475 480  
 Asn Gly Gly Val Thr Val Ser Ser Ile Ser Leu Pro Phe Phe Lys Lys  
 485 490 495  
 Phe Asp Ser Ser Ala Thr Ser Gly Lys Lys Tyr Thr Val Gly Thr Ser  
 500 505 510  
 Asp Phe Asn Asn Leu Ala Gln Asn Ile Ala Leu Ala Ala Asp Arg Phe  
 515 520 525  
 Leu Ser Thr Val Gln Leu His Ala His Asn Asn Gly Ser Leu Ala Glu  
 530 535 540  
 Glu Phe Asp Arg Thr Thr Gly Leu Ser Thr Gly Ala Arg Asp Leu Thr  
 545 550 555 560  
 Trp Ser His Ala Ser Leu Ile Thr Ala Ser Tyr Ala Lys Ala Gly Ala  
 565 570 575  
 Pro Ala Ala

&lt;210&gt; 50

&lt;211&gt; 1737

&lt;212&gt; DNA

<213> *Rhizopus oryzae*

&lt;400&gt; 50

gcctccatcc cgctctccgc ctccgtgcag ctcgactcct acaactacga cggctccacc 60  
 ttctccggca aaatctacgt gaagaacatc gcctactcca agaagggtgac cgtgatctac 120  
 gccgacggct ccgacaactg gaacaacaac ggcaacacca tcgccgctc ctactccgcc 180  
 ccgatctccg gctccaacta cgagtactgg accttctccg cctccatcaa cggcatcaag 240  
 gagttctaca tcaagtacga ggtgtccggc aagacctact acgacaacaa caactccgcc 300  
 aactaccagg tgtccacctc caagccgacc accaccaccg ccaccgccac caccaccacc 360

```

gccccgtcca cctccaccac caccgcgccc tcccgtccg agccggccac ctccccgacc 420
ggcaactcca ccatctcctc ctggatcaag aagcaggagg gcatctcccc ctccgccatg 480
ctccgcaaca tcaacccgcc gggtccgcc accggcttca tcgccgcctc cctctccacc 540
gccggcccg actactacta cgcttgacc cgcgacgcc ccctcacctc caacgtgatc 600
gtgtacgagt acaacaccac cctctccggc aacaagacca tcctcaacgt gctcaaggac 660
tacgtgacct tctccgtgaa gaccagtc acctccaccg tgtgcaactg cctcggcgag 720
ccgaagttca acccggacgc ctccggctac accggcgctt gggggcgccc gcagaacgac 780
ggcccgccg agcgcgccac caccttcac cttctcgccg actcctacct caccagacc 840
aaggacgcct cctacgtgac cggcacctc aagccggcca tcttcaagga cctcgactac 900
gtggtgaacg tgtggtccaa cggtgcttc gacctctggg aggaggtgaa cggcggtgcac 960
ttctacacc tcattggtgat gcgcaaggc ctcctcctc ggcgcgactt cgccaagcgc 1020
aacggcgact ccacccgcgc ctccacctac tcctccaccg cctccaccat cgccaacaaa 1080
atctcctcct tctgggtgtc ctccaacaac tggatacagg tgtcccagtc cgtgaccggc 1140
ggcgtgtcca agaagggcct cgacgtgtcc accctcctc cgccaacct cggctccgtg 1200
gacgacggct tcttcacccc gggctccgag aagatcctc ccaccgcctt ggccgtggag 1260
gactccttcg cctccctcta cccgatcaac aagaacctc cgtcctacct cggcaactcc 1320
atcggcgctt acccgaggga cactacaac ggcaacggca actcccagg caactcctgg 1380
ttctcgccc tgaccggcta cgccgagctg tactaccgcg ccatcaagga gtggatcggc 1440
aacggcggtg tgaccgtgtc ctccatctcc ctcccgttct tcaagaagtt cgactcctcc 1500
gccacctccg gcaagaagta caccgtgggc acctccgact tcaacaacct cgcccagaac 1560
atcgccctcg ccgcccaccc cttcctctcc accgtgcagc tccacgcccc caacaacggc 1620
tccctcgccg aggagttcga ccgcaccacc ggctctcca ccggcgccc cgacctcacc 1680
tggccccacg cctccctcat caccgcctcc tacgccaagg ccggcgcccc ggccgcc 1737

```

<210> 51

<211> 439

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic

<400> 51

```

Met Ala Lys His Leu Ala Ala Met Cys Trp Cys Ser Leu Leu Val Leu
1          5          10          15
Val Leu Leu Cys Leu Gly Ser Gln Leu Ala Gln Ser Gln Val Leu Phe
20          25          30
Gln Gly Phe Asn Trp Glu Ser Trp Lys Lys Gln Gly Gly Trp Tyr Asn
35          40          45
Tyr Leu Leu Gly Arg Val Asp Asp Ile Ala Ala Thr Gly Ala Thr His
50          55          60
Val Trp Leu Pro Gln Pro Ser His Ser Val Ala Pro Gln Gly Tyr Met
65          70          75          80
Pro Gly Arg Leu Tyr Asp Leu Asp Ala Ser Lys Tyr Gly Thr His Ala
85          90          95
Glu Leu Lys Ser Leu Thr Ala Ala Phe His Ala Lys Gly Val Gln Cys
100          105          110
Val Ala Asp Val Val Ile Asn His Arg Cys Ala Asp Tyr Lys Asp Gly
115          120          125
Arg Gly Ile Tyr Cys Val Phe Glu Gly Gly Thr Pro Asp Ser Arg Leu
130          135          140
Asp Trp Gly Pro Asp Met Ile Cys Ser Asp Asp Thr Gln Tyr Ser Asn
145          150          155          160
Gly Arg Gly His Arg Asp Thr Gly Ala Asp Phe Ala Ala Ala Pro Asp

```

```

      165      170      175
Ile Asp His Leu Asn Pro Arg Val Gln Gln Glu Leu Ser Asp Trp Leu
      180      185      190
Asn Trp Leu Lys Ser Asp Leu Gly Phe Asp Gly Trp Arg Leu Asp Phe
      195      200      205
Ala Lys Gly Tyr Ser Ala Ala Val Ala Lys Val Tyr Val Asp Ser Thr
      210      215      220
Ala Pro Thr Phe Val Val Ala Glu Ile Trp Ser Ser Leu His Tyr Asp
      225      230      235
Gly Asn Gly Glu Pro Ser Ser Asn Gln Asp Ala Asp Arg Gln Glu Leu
      245      250      255
Val Asn Trp Ala Gln Ala Val Gly Gly Pro Ala Ala Ala Phe Asp Phe
      260      265      270
Thr Thr Lys Gly Val Leu Gln Ala Ala Val Gln Gly Glu Leu Trp Arg
      275      280      285
Met Lys Asp Gly Asn Gly Lys Ala Pro Gly Met Ile Gly Trp Leu Pro
      290      295      300
Glu Lys Ala Val Thr Phe Val Asp Asn His Asp Thr Gly Ser Thr Gln
      305      310      315
Asn Ser Trp Pro Phe Pro Ser Asp Lys Val Met Gln Gly Tyr Ala Tyr
      325      330      335
Ile Leu Thr His Pro Gly Thr Pro Cys Ile Phe Tyr Asp His Val Phe
      340      345      350
Asp Trp Asn Leu Lys Gln Glu Ile Ser Ala Leu Ser Ala Val Arg Ser
      355      360      365
Arg Asn Gly Ile His Pro Gly Ser Glu Leu Asn Ile Leu Ala Ala Asp
      370      375      380
Gly Asp Leu Tyr Val Ala Lys Ile Asp Asp Lys Val Ile Val Lys Ile
      385      390      395
Gly Ser Arg Tyr Asp Val Gly Asn Leu Ile Pro Ser Asp Phe His Ala
      405      410      415
Val Ala His Gly Asn Asn Tyr Cys Val Trp Glu Lys His Gly Leu Arg
      420      425      430
Val Pro Ala Gly Arg His His
      435

```

&lt;210&gt; 52

&lt;211&gt; 1320

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; synthetic

&lt;400&gt; 52

```

atggcgaagc acttggtgctgc catgtgctgg tgcagcctcc tagtgcttgt actgctctgc 60
ttgggctccc agctggccca atcccaggtc ctcttcagg ggttcaactg ggagtcgtgg 120
aagaagcaag gtgggtggtgta caactacctc ctggggcggg tggacgacat cgccgcgacg 180
ggggccacgc acgtctggct cccgcagccg tcgcactcgg tggcgccgca ggggtacatg 240
ccccggcggc tctacgacct ggacgcgtcc aagtacggca cccacgcgga gctcaagtcg 300
ctcaccgcgg cgttcacacgc caagggcgtc cagtgcgtcg ccgacgtcgt gatcaaccac 360
cgctgcgccg actacaagga cggccgcggc atctactgcg tcttcgaggg cggcacgccc 420
gacagccgcc tcgactgggg ccccgacatg atctgcagcg acgacacgca gtactccaac 480

```

```

gggcgcgggc accgcgacac gggggccgac ttcgccgccg cggccgacat cgaccacctc 540
aaccgcgcgc tgcagcagga gctctcggac tggctcaact ggctcaagtc cgacctcggc 600
ttcgacggct ggccgctcga cttcgccaag ggctactccg ccgccgtcgc caaggtgtac 660
gtcgacagca ccgccccac cttcgtcgtc gccgagatat ggagctccct ccactacgac 720
ggcaacggcg agccgtccag caaccaggac gccgacaggc aggagctggt caactgggcg 780
caggcggtgg gcggccccgc cgcggcggtc gacttcacca ccaagggcgt gctgcaggcg 840
gccgtccagg gcgagctgtg gcgcatgaag gacggcaacg gcaaggcgcc cgggatgatc 900
ggctggctgc cggagaaggc cgtcacgttc gtcgacaacc acgacaccgg ctccacgcag 960
aactcgtggc cattccccctc cgacaaggte atgcagggtc acgcctatat cctcacgcac 1020
ccaggaactc catgcattct ctacgaccac gttttcgact ggaacctgaa gcaggagatc 1080
agcgcgctgt ctgcgggtgag gtcaagaaac gggatccacc cggggagcga gctgaacatc 1140
ctcgccgccg acggggatct ctacgtcgcc aagattgacg acaaggatcat cgtgaagatc 1200
gggtcacggc acgacgtcgg gaacctgatc ccctcagact tccacgccgt tgcccctggc 1260
aacaactact gcgtttggga gaagcacggt ctgagagttc cagcggggcg gcaccactag 1320

```

<210> 53

<211> 45

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic

<400> 53

```

Ala Thr Gly Gly Thr Thr Thr Ala Thr Thr Thr Gly Ser Gly Gly
 1           5           10           15
Val Thr Ser Thr Ser Lys Thr Thr Thr Ala Ser Lys Thr Ser Thr
      20           25           30
Thr Thr Ser Ser Thr Ser Cys Thr Thr Pro Thr Ala Val
      35           40           45

```

<210> 54

<211> 137

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic

<400> 54

```

gccaccggcg gcaccaccac caccgccacc accaccggct ccggcggcgt gacctccacc 60
tccaagacca ccaccaccgc ctccaagacc tccaccacca cctcctccac ctccctgcacc 120
accccgaccg ccgtgtc                                     137

```

<210> 55

<211> 300

<212> PRT

<213> *Pyrococcus furiosus*

<400> 55

```

Ile Tyr Phe Val Glu Lys Tyr His Thr Ser Glu Asp Lys Ser Thr Ser
 1           5           10           15
Asn Thr Ser Ser Thr Pro Pro Gln Thr Thr Leu Ser Thr Thr Lys Val

```

	20		25		30										
Leu	Lys	Ile	Arg	Tyr	Pro	Asp	Asp	Gly	Glu	Trp	Pro	Gly	Ala	Pro	Ile
	35					40					45				
Asp	Lys	Asp	Gly	Asp	Gly	Asn	Pro	Glu	Phe	Tyr	Ile	Glu	Ile	Asn	Leu
	50					55					60				
Trp	Asn	Ile	Leu	Asn	Ala	Thr	Gly	Phe	Ala	Glu	Met	Thr	Tyr	Asn	Leu
	65				70				75					80	
Thr	Ser	Gly	Val	Leu	His	Tyr	Val	Gln	Gln	Leu	Asp	Asn	Ile	Val	Leu
			85					90					95		
Arg	Asp	Arg	Ser	Asn	Trp	Val	His	Gly	Tyr	Pro	Glu	Ile	Phe	Tyr	Gly
			100					105					110		
Asn	Lys	Pro	Trp	Asn	Ala	Asn	Tyr	Ala	Thr	Asp	Gly	Pro	Ile	Pro	Leu
		115				120					125				
Pro	Ser	Lys	Val	Ser	Asn	Leu	Thr	Asp	Phe	Tyr	Leu	Thr	Ile	Ser	Tyr
	130					135					140				
Lys	Leu	Glu	Pro	Lys	Asn	Gly	Leu	Pro	Ile	Asn	Phe	Ala	Ile	Glu	Ser
	145				150				155					160	
Trp	Leu	Thr	Arg	Glu	Ala	Trp	Arg	Thr	Thr	Gly	Ile	Asn	Ser	Asp	Glu
			165					170						175	
Gln	Glu	Val	Met	Ile	Trp	Ile	Tyr	Tyr	Asp	Gly	Leu	Gln	Pro	Ala	Gly
		180						185					190		
Ser	Lys	Val	Lys	Glu	Ile	Val	Val	Pro	Ile	Ile	Val	Asn	Gly	Thr	Pro
	195					200					205				
Val	Asn	Ala	Thr	Phe	Glu	Val	Trp	Lys	Ala	Asn	Ile	Gly	Trp	Glu	Tyr
	210					215					220				
Val	Ala	Phe	Arg	Ile	Lys	Thr	Pro	Ile	Lys	Glu	Gly	Thr	Val	Thr	Ile
	225				230					235				240	
Pro	Tyr	Gly	Ala	Phe	Ile	Ser	Val	Ala	Ala	Asn	Ile	Ser	Ser	Leu	Pro
			245					250						255	
Asn	Tyr	Thr	Glu	Leu	Tyr	Leu	Glu	Asp	Val	Glu	Ile	Gly	Thr	Glu	Phe
		260				265						270			
Gly	Thr	Pro	Ser	Thr	Thr	Ser	Ala	His	Leu	Glu	Trp	Trp	Ile	Thr	Asn
	275					280						285			
Ile	Thr	Leu	Thr	Pro	Leu	Asp	Arg	Pro	Leu	Ile	Ser				
	290					295					300				

&lt;210&gt; 56

&lt;211&gt; 903

&lt;212&gt; DNA

<213> *Pyrococcus furiosus*

&lt;400&gt; 56

```

atctacttcg tggagaagta ccacacctcc gaggacaagt ccacctccaa cacctcctcc 60
accccgccgc agaccaccct ctccaccacc aaggtgctca agatccgcta cccggacgac 120
ggcgagtggc ccggcgcccc gatcgacaag gacggcgacg gcaaccggga gttctacatc 180
gagatcaacc tctggaacat cctcaacgcc accggcttcg ccgagatgac ctacaacctc 240
actagtggcg tgctccacta cgtgcagcag ctcgacaaca tcgtgctccg cgaccgctcc 300
aactgggtgc acggctaccc ggaaatcttc tacggcaaca agccgtggaa cgccaactac 360
gccaccgacg gccgatccc gctcccgtcc aaggtgtcca acctcaccga cttctacctc 420
accatctcct acaagctcga gccgaagaac ggtctcccga tcaacttcgc catcgagtcc 480
tggtcacccc gcgaggcctg gcgcaccacc ggcacaaact ccgacgagca ggaggtgatg 540
atctggatct actacgacgg cctccagccc gcgggctcca aggtgaagga gatcgtggtg 600
ccgatcatcg tgaacggcac cccggtgaac gccaccttcg aggtgtggaa ggccaacatc 660
ggctgggagt acgtggcctt ccgcatcaag accccgatca aggagggcac cgtgaccatc 720

```



<210> 57  
<211> 387  
<212> PRT  
<213> Thermus flavus

229

	340		345		350
Ala	Leu Lys Arg Ala Glu Leu Pro	Leu Glu Thr Lys Arg Arg Arg Gly			
	355		360		365
Tyr	Ala Leu Glu Arg Leu Asp Gln	Leu Ala Val Glu Tyr Leu Leu Gly			
	370		375		380
Val	Arg Gly				
385					

<210> 58  
 <211> 978  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 58  
 atggggaaga acggcaacct gtgctgcttc tctctgctgc tgctttcttct cgccggggtg 60  
 gcgtccggcc atcaaatcta ctctgtggag aagtaccaca cctccgagga caagtccacc 120  
 tccaacacct cctccacccc gccgcagacc accctctcca ccaccaaggt gctcaagatc 180  
 cgctacccgg acgacgggtga gtggcccggc gccccgatcg acaaggacgg cgacggcaac 240  
 ccggagtctt acatcgagat caacctctgg aacatcctca acgccaccgg ctctgccgag 300  
 atgacctaca acctcactag tggcgtgctc cactacgtgc agcagctcga caacatcgtg 360  
 ctccgcgacc gctccaactg ggtgcacggc tacccgaaa tcttctacgg caacaagccg 420  
 tggaacgcca actacgccac cgacggcccg atcccgtcc cgtccaaggt gtccaacctc 480  
 accgacttct acctcaccat ctccctacaag ctcgagccga agaacgggtct cccgatcaac 540  
 ttccgcatcg agtccctggc caccgcgag gcctggcgca ccaccggcat caactccgac 600  
 gagcaggagg tgatgatctg gatctactac gacggcctcc agcccgcggg ctccaagggtg 660  
 aaggagatcg tggtgccgat catcgtgaac ggcaccccg tgaacgccac ctctgagggtg 720  
 tggaaggcca acatcggtg ggagtacgtg gccttcgcga tcaagacccc gatcaaggag 780  
 ggcacgtga ccatcccgta cggcgccctc atctccgtgg ccgccaacat ctctccctc 840  
 ccgaactaca ccgagaagta cctcgaggac gtggagatcg gcaccgaggt cggcaccctc 900  
 tccaccacct ccgcccacct cgagtgggtg atcaccaaca tcacctcac cccgctcgac 960  
 cgcccgctca tctcctag 978

<210> 59  
 <211> 1920  
 <212> DNA  
 <213> *Aspergillus niger*

<400> 59  
 atgtccttcc gtcctctcct cgccctctcc ggctcgtgt gcaccggcct cgccaacgtg 60  
 atctccaagc gcgccaccct cgactcctgg ctctccaacg aggccaccgt ggcccgcacc 120  
 gccatcctca acaacatcgg cgccgacggc gcctgggtgt ccggcgccga ctccggcatc 180  
 gtggtggcct ccccgctccac cgacaaccgg gactacttct acacctggac ccgcgactcc 240  
 ggctcgtgc tcaagaccct cgtggacctc ttccgcaacg gcgacacctc cctcctctcc 300  
 accatcgaga actacatctc cgcccaggcc atcgtgcagg gcatctccaa cccgtccggc 360  
 gacctctcct ccggcgccgg cctcggcgag ccgaagttca acgtggacga gaccgcctac 420

```

accggctcct ggggcccggc gcagcgcgac ggcccggccc tccgcgccac cgccatgac 480
ggcttcggcc agtggctcct cgacaacggc tacacctcca ccgccaccga catcgtgtgg 540
ccgctcgtgc gcaacgacct ctctacgtg gccagtagt ggaaccagac cggctacgac 600
ctctgggagg aggtgaacgg ctctccttc ttaccatcg ccgtgcagca ccgcgccctc 660
5 gtggagggct ccgccttcgc caccgcctg ggctcctcct gctcctggg cgactcccag 720
gccccggaga tctctgcta cctccagtc ttctggacc gctccttcat cctcgccaac 780
ttcgactcct ccgcctccgg caaggacgcc aacaccctcc tcggctccat ccacaccttc 840
gaccggagg ccgcctgcga cgactccacc ttccagcgt gctccccgg cgccctcgcc 900
aaccacaagg aggtggtgga ctcttcgcg tccatctaca cctcaacga cggcctctcc 960
10 gactccgagg ccgtggcgt ggcccgctac ccggaggaca cctactaaa cggcaaccgg 1020
tggttcctct gcacctcgc cgcgcgcgag cagctctac agccctcta ccagtgggac 1080
aagcagggt cctcgagggt gaccgacgt tccctcgact tctcaaggc cctctactcc 1140
gacgcggcca ccggcaccta ctctcctcc tctccacct actcctccat cgtggacgcc 1200
gtgaagacct tcgccgacgg ctctggtgc atcgtggaga cccacgcgc ctccaacggc 1260
15 tccatgtccg agcagtacga caagtccgac ggcgagcag tctccgccc cgacctcacc 1320
tggtcctacg ccgcctcct caccgccaac aaccgcccga actccgtggt gccggcctcc 1380
tggggcgaga cctccgcctc ctccgtgccc ggacactgc ccgccacct cgccatcgcc 1440
acctactcct ccgtgaccgt gacctcctgg ccgtccatcg tggccaccgg cggcaccacc 1500
accaccgcca ccccgaccgg ctccggctcc gtgacctcca cctccaagac caccgccacc 1560
20 gcctccaaga cctccacctc cactcctcc acctcctgca ccaccccgac cgcctggcc 1620
gtgaccttcg acctcaccgc caccaccacc tacggcgaga acatctacct cgtgggctcc 1680
atctcccagc tcggcgactg ggagacctcc gacggcatcg cctctccgc cgacaagtac 1740
acctcctccg acccgctctg gtacgtgacc gtgacctcc cggccggcga gtccttcgag 1800
tacaagttca tccgcatcga gtccgacgac tccgtggagt gggagtccga cccgaaccgc 1860
25 gagtacaccg tgccgcaggc ctgcggcacc tccaccgcca ccgtgaccga cacctggcgc 1920

```

<210> 60

<211> 6

30 <212> PRT

<213> Artificial Sequence

<220>

<223> synthetic

35

<400> 60

Ser Glu Lys Asp Glu Leu

1

5

40

<210> 61

<211> 561

<212> DNA

45 <213> Artificial Sequence

<220>

<223> Xylanase BD7436

50

<220>

<221> CDS

<222> (1)..(561)

55

<400> 61

atg gct agc acc ttc tac tgg cat ttg tgg acc gac ggc atc ggc acc  
Met Ala Ser Thr Phe Tyr Trp His Leu Trp Thr Asp Gly Ile Gly Thr

48

1	5	10	15	
gtg aac gct acc aac ggc agc gac ggc aac tac agc gtg agc tgg agc				96
Val Asn Ala Thr Asn Gly Ser Asp Gly Asn Tyr Ser Val Ser Trp Ser				
5	20	25	30	
aac tgc ggc aac ttc gtg gtg ggc aag ggc tgg acc acc ggc agc gct				144
Asn Cys Gly Asn Phe Val Val Gly Lys Gly Trp Thr Thr Gly Ser Ala				
10	35	40	45	
acc agg gtg atc aac tac aac gct cat gct ttc agc gtg gtg ggc aac				192
Thr Arg Val Ile Asn Tyr Asn Ala His Ala Phe Ser Val Val Gly Asn				
15	50	55	60	
gct tac ttg gct ttg tac ggc tgg acc agg aac agc ttg atc gag tac				240
Ala Tyr Leu Ala Leu Tyr Gly Trp Thr Arg Asn Ser Leu Ile Glu Tyr				
20	65	70	75	
tac gtg gtg gac agc tgg ggc acc tac agg cca acc ggc acc tac aag				288
Tyr Val Val Asp Ser Trp Gly Thr Tyr Arg Pro Thr Gly Thr Tyr Lys				
25	85	90	95	
ggc acc gtg acc agc gac ggc ggc acc tac gac atc tac acc acc acc				336
Gly Thr Val Thr Ser Asp Gly Gly Thr Tyr Asp Ile Tyr Thr Thr Thr				
30	100	105	110	
agg acc aac gct cca agc atc gac ggc aac aac acc acc ttc acc caa				384
Arg Thr Asn Ala Pro Ser Ile Asp Gly Asn Asn Thr Thr Phe Thr Gln				
35	115	120	125	
ttc tgg agc gtg agg caa agc aag agg cca atc ggc acc aac aac acc				432
Phe Trp Ser Val Arg Gln Ser Lys Arg Pro Ile Gly Thr Asn Asn Thr				
40	130	135	140	
atc acc ttc agc aac cat gtg aac gct tgg aag agc aag ggc atg aac				480
Ile Thr Phe Ser Asn His Val Asn Ala Trp Lys Ser Lys Gly Met Asn				
45	145	150	155	
ttg ggc agc agc tgg agc tac caa gtg ttg gct acc gag ggc tac caa				528
Leu Gly Ser Ser Trp Ser Tyr Gln Val Leu Ala Thr Glu Gly Tyr Gln				
50	165	170	175	
agc agc ggc tac agc aac gtg acc gtg tgg tag				561
Ser Ser Gly Tyr Ser Asn Val Thr Val Trp				
55	180	185		
<210> 62				
<211> 186				
<212> PRT				
<213> Artificial Sequence				
<220>				
<223> Synthetic Construct				
<400> 62				

Met Ala Ser Thr Phe Tyr Trp His Leu Trp Thr Asp Gly Ile Gly Thr  
 1 5 10 15  
 5 Val Asn Ala Thr Asn Gly Ser Asp Gly Asn Tyr Ser Val Ser Trp Ser  
 20 25 30  
 10 Asn Cys Gly Asn Phe Val Val Gly Lys Gly Trp Thr Thr Gly Ser Ala  
 35 40 45  
 Thr Arg Val Ile Asn Tyr Asn Ala His Ala Phe Ser Val Val Gly Asn  
 50 55 60  
 15 Ala Tyr Leu Ala Leu Tyr Gly Trp Thr Arg Asn Ser Leu Ile Glu Tyr  
 65 70 75 80  
 20 Tyr Val Val Asp Ser Trp Gly Thr Tyr Arg Pro Thr Gly Thr Tyr Lys  
 85 90 95  
 25 Gly Thr Val Thr Ser Asp Gly Gly Thr Tyr Asp Ile Tyr Thr Thr Thr  
 100 105 110  
 30 Arg Thr Asn Ala Pro Ser Ile Asp Gly Asn Asn Thr Thr Phe Thr Gln  
 115 120 125  
 Phe Trp Ser Val Arg Gln Ser Lys Arg Pro Ile Gly Thr Asn Asn Thr  
 130 135 140  
 35 Ile Thr Phe Ser Asn His Val Asn Ala Trp Lys Ser Lys Gly Met Asn  
 145 150 155 160  
 40 Leu Gly Ser Ser Trp Ser Tyr Gln Val Leu Ala Thr Glu Gly Tyr Gln  
 165 170 175  
 45 Ser Ser Gly Tyr Ser Asn Val Thr Val Trp  
 180 185  
 <210> 63  
 50 <211> 561  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 55 <223> Xylanase BD6002A

```

<220>
<221> CDS
<222> (1)..(561)

5  <400> 63
   atg gct agc acc gac tac tgg caa aac tgg acc gac ggc ggc ggc acc    48
   Met Ala Ser Thr Asp Tyr Trp Gln Asn Trp Thr Asp Gly Gly Gly Thr
   1                               5                               10                               15

10  gtg aac gct acc aac ggc agc gac ggc aac tac agc gtg agc tgg agc    96
   Val Asn Ala Thr Asn Gly Ser Asp Gly Asn Tyr Ser Val Ser Trp Ser
                               20                               25                               30

15  aac tgc ggc aac ttc gtg gtg ggc aag ggc tgg acc acc ggc agc gct    144
   Asn Cys Gly Asn Phe Val Val Gly Lys Gly Trp Thr Thr Gly Ser Ala
                               35                               40                               45

20  acc agg gtg atc aac tac aac gct ggc gct ttc agc cca agc ggc aac    192
   Thr Arg Val Ile Asn Tyr Asn Ala Gly Ala Phe Ser Pro Ser Gly Asn
                               50                               55                               60

25  ggc tac ttg gct ttg tac ggc tgg acc agg aac agc ttg atc gag tac    240
   Gly Tyr Leu Ala Leu Tyr Gly Trp Thr Arg Asn Ser Leu Ile Glu Tyr
   65                               70                               75                               80

   tac gtg gtg gac agc tgg ggc acc tac agg cca acc ggc acc tac aag    288
   Tyr Val Val Asp Ser Trp Gly Thr Tyr Arg Pro Thr Gly Thr Tyr Lys
                               85                               90                               95

30  ggc acc gtg acc agc gac ggc ggc acc tac gac atc tac acc acc acc    336
   Gly Thr Val Thr Ser Asp Gly Gly Thr Tyr Asp Ile Tyr Thr Thr Thr
                               100                              105                              110

35  agg acc aac gct cca agc atc gac ggc aac aac acc acc ttc acc caa    384
   Arg Thr Asn Ala Pro Ser Ile Asp Gly Asn Asn Thr Thr Phe Thr Gln
                               115                              120                              125

40  ttc tgg agc gtg agg caa agc aag agg cca atc ggc acc aac aac acc    432
   Phe Trp Ser Val Arg Gln Ser Lys Arg Pro Ile Gly Thr Asn Asn Thr
                               130                              135                              140

   atc acc ttc agc aac cat gtg aac gct tgg aag agc aag ggc atg aac    480
   Ile Thr Phe Ser Asn His Val Asn Ala Trp Lys Ser Lys Gly Met Asn
   145                              150                              155                              160

45  ttg ggc agc agc tgg agc tac caa gtg ttg gct acc gag ggc tac caa    528
   Leu Gly Ser Ser Trp Ser Tyr Gln Val Leu Ala Thr Glu Gly Tyr Gln
                               165                              170                              175

50  agc agc ggc tac agc aac gtg acc gtg tgg tag    561
   Ser Ser Gly Tyr Ser Asn Val Thr Val Trp
                               180                              185

55  <210> 64
   <211> 186
   <212> PRT

```

<213> Artificial Sequence

<220>

<223> Synthetic Construct

5

<400> 64

Met Ala Ser Thr Asp Tyr Trp Gln Asn Trp Thr Asp Gly Gly Gly Thr  
1 5 10 15

Val Asn Ala Thr Asn Gly Ser Asp Gly Asn Tyr Ser Val Ser Trp Ser  
20 25 30

Asn Cys Gly Asn Phe Val Val Gly Lys Gly Trp Thr Thr Gly Ser Ala  
35 40 45

Thr Arg Val Ile Asn Tyr Asn Ala Gly Ala Phe Ser Pro Ser Gly Asn  
50 55 60

Gly Tyr Leu Ala Leu Tyr Gly Trp Thr Arg Asn Ser Leu Ile Glu Tyr  
65 70 75 80

Tyr Val Val Asp Ser Trp Gly Thr Tyr Arg Pro Thr Gly Thr Tyr Lys  
85 90 95

Gly Thr Val Thr Ser Asp Gly Gly Thr Tyr Asp Ile Tyr Thr Thr Thr  
100 105 110

Arg Thr Asn Ala Pro Ser Ile Asp Gly Asn Asn Thr Thr Phe Thr Gln  
115 120 125

Phe Trp Ser Val Arg Gln Ser Lys Arg Pro Ile Gly Thr Asn Asn Thr  
130 135 140

Ile Thr Phe Ser Asn His Val Asn Ala Trp Lys Ser Lys Gly Met Asn  
145 150 155 160

Leu Gly Ser Ser Trp Ser Tyr Gln Val Leu Ala Thr Glu Gly Tyr Gln  
165 170 175

Ser Ser Gly Tyr Ser Asn Val Thr Val Trp  
180 185

55

<210> 65

<211> 561

```

<212> DNA
<213> Artificial Sequence

<220>
5 <223> Xylanase BD6002B

<220>
<221> CDS
10 <222> (1)..(561)

<400> 65
atg gcc tcc acc gac tac tgg cag aac tgg acc gac ggc ggc ggc acc 48
Met Ala Ser Thr Asp Tyr Trp Gln Asn Trp Thr Asp Gly Gly Gly Thr
15 1 5 10 15

gtg aac gcc acc aac ggc tcc gac ggc aac tac tcc gtg tcc tgg tcc 96
Val Asn Ala Thr Asn Gly Ser Asp Gly Asn Tyr Ser Val Ser Trp Ser
20 20 25 30

aac tgc ggc aac ttc gtg gtg ggc aag ggc tgg acc acc ggc tcc gcc 144
Asn Cys Gly Asn Phe Val Val Gly Lys Gly Trp Thr Thr Gly Ser Ala
35 40 45

acc cgc gtg atc aac tac aac gcc ggc gcc ttc tcc ccg tcc ggc aac 192
Thr Arg Val Ile Asn Tyr Asn Ala Gly Ala Phe Ser Pro Ser Gly Asn
50 55 60

ggc tac ctc gcc ctc tac ggc tgg acc cgc aac tcc ctc atc gag tac 240
Gly Tyr Leu Ala Leu Tyr Gly Trp Thr Arg Asn Ser Leu Ile Glu Tyr
65 70 75 80

tac gtg gtg gac tcc tgg ggc acc tac cgc ccg acc ggc acc tac aag 288
Tyr Val Val Asp Ser Trp Gly Thr Tyr Arg Pro Thr Gly Thr Tyr Lys
35 85 90 95

ggc acc gtg acc tcc gac ggc ggc acc tac gac atc tac acc acc acc 336
Gly Thr Val Thr Ser Asp Gly Gly Thr Tyr Asp Ile Tyr Thr Thr Thr
40 100 105 110

cgc acc aac gcc ccg tcc atc gac ggc aac aac acc acc ttc acc cag 384
Arg Thr Asn Ala Pro Ser Ile Asp Gly Asn Asn Thr Thr Phe Thr Gln
115 120 125

ttc tgg tcc gtg cgc cag tcc aag cgc ccg atc ggc acc aac aac acc 432
Phe Trp Ser Val Arg Gln Ser Lys Arg Pro Ile Gly Thr Asn Asn Thr
130 135 140

atc acc ttc tcc aac cac gtg aac gcc tgg aag tcc aag ggc atg aac 480
Ile Thr Phe Ser Asn His Val Asn Ala Trp Lys Ser Lys Gly Met Asn
145 150 155 160

ctc ggc tcc tcc tgg tcc tac cag gtg ctc gcc acc gag ggc tac cag 528
Leu Gly Ser Ser Trp Ser Tyr Gln Val Leu Ala Thr Glu Gly Tyr Gln
55 165 170 175

tcc tcc ggc tac tcc aac gtg acc gtg tgg tga 561

```



Ser Ser Gly Tyr Ser Asn Val Thr Val Trp  
 180 185

5 <210> 66  
 <211> 186  
 <212> PRT  
 <213> Artificial Sequence

10 <220>  
 <223> Synthetic Construct  
 <400> 66

15 Met Ala Ser Thr Asp Tyr Trp Gln Asn Trp Thr Asp Gly Gly Gly Thr  
 1 5 10 15

20 Val Asn Ala Thr Asn Gly Ser Asp Gly Asn Tyr Ser Val Ser Trp Ser  
 20 25 30

25 Asn Cys Gly Asn Phe Val Val Gly Lys Gly Trp Thr Thr Gly Ser Ala  
 35 40 45

Thr Arg Val Ile Asn Tyr Asn Ala Gly Ala Phe Ser Pro Ser Gly Asn  
 50 55 60

30 Gly Tyr Leu Ala Leu Tyr Gly Trp Thr Arg Asn Ser Leu Ile Glu Tyr  
 65 70 75 80

35 Tyr Val Val Asp Ser Trp Gly Thr Tyr Arg Pro Thr Gly Thr Tyr Lys  
 85 90 95

40 Gly Thr Val Thr Ser Asp Gly Gly Thr Tyr Asp Ile Tyr Thr Thr Thr  
 100 105 110

45 Arg Thr Asn Ala Pro Ser Ile Asp Gly Asn Asn Thr Thr Phe Thr Gln  
 115 120 125

Phe Trp Ser Val Arg Gln Ser Lys Arg Pro Ile Gly Thr Asn Asn Thr  
 130 135 140

50 Ile Thr Phe Ser Asn His Val Asn Ala Trp Lys Ser Lys Gly Met Asn  
 145 150 155 160

55 Leu Gly Ser Ser Trp Ser Tyr Gln Val Leu Ala Thr Glu Gly Tyr Gln  
 165 170 175

Ser Ser Gly Tyr Ser Asn Val Thr Val Trp  
 180 185

5  
 <210> 67  
 <211> 2071  
 <212> DNA  
 <213> Oryza sativa

10  
 <220>  
 <221> misc\_feature  
 <222> (1)..(2071)  
 15 <223> Promoter

<400> 67  
 tccatgctgt cctactactt gcttcatccc cttctacatt ttgttctggt ttttggcctg 60  
 20 catttcggat catgatgtat gtgatttcca atctgctgca atatgaatgg agactctgtg 120  
 ctaaccatca acaacatgaa atgcttatga ggcctttgct gagcagccaa tcttgccctgt 180  
 gtttatgtct tcacaggccg aattcctctg ttttgttttt caccctcaat atttggaac 240  
 25 atttatctag gttgtttgtg tccaggccca taaatcacat atgatgttgt cgtattggat 300  
 gtgaatgtgg tggcgtgttc agtgccttgg atttgagttt gatgagagtt gcttctgggt 360  
 30 caccactcac cattatcgat gctcctcttc agcataagggt aaaagtcttc cctgtttacg 420  
 ttattttacc cactatgggt gcttggttg gtttttctt gattgcttat gccatggaaa 480  
 gtcatttgat atgttgaact tgaattaact gtagaattgt atacatgttc catttggtt 540  
 35 gtacttcctt cttttctatt agtagcctca gatgagtgtg aaaaaaacag attatataac 600  
 ttgccctata aatcatttga aaaaaatatt gtacagttag aaattgatat atagtgaatt 660  
 40 tttaagagca tgttttccta aagaagtata tattttctat gtacaaaggc cattgaagta 720  
 attgtagata caggataatg tagacttttt ggacttacac tgctacctt aagtaacaat 780  
 catgagcaat agtgttgcaa tgatatttag gctgcattcg tttactctct tgatttccat 840  
 45 gagcacgctt cccaaactgt taaactctgt gttttttgcc aaaaaaaaaat gcataggaaa 900  
 gttgctttta aaaaatcata tcaatccatt ttttaagtta tagctaatac ttaattaatc 960  
 50 atgcgcta atagtcactct gtttttcgta ctagagagat tgttttgaac cagcactcaa 1020  
 gaacacagcc ttaaccagc caaataatgc tacaacctac cagtccacac ctcttgtaaa 1080  
 gcatttggtg catggaaaag ctaagatgac agcaacctgt tcaggaaaac aactgacaag 1140  
 55 gtcatagga gagggagctt ttggaaagggt gccgtgcagt tcaaacaatt agttagcagt 1200

aggggtgttg tttttgctca cagcaataag aagttaatca tgggtgtaggc aacccaaata 1260  
 aaacacccaaa atatgcacaa ggcagtttgt tgtattctgt agtacagaca aaactaaaag 1320  
 5 taatgaaaga agatgtggtg ttagaaaagg aaacaatatc atgagtaatg tgtgggcatt 1380  
 atggggaccac gaaataaaaa gaacattttg atgagtcgtg tatectcgat gagcctcaaa 1440  
 10 agttctctca ccccgataa gaaaccctta agcaatgtgc aaagtttgca ttctccactg 1500  
 acataatgca aaataagata tcatcgatga catagcaact catgcatcat atcatgcctc 1560  
 tctcaaccta ttcattccta ctcatctaca taagtatctt cagctaaatg ttagaacata 1620  
 15 aacccataag tcacgtttga tgagtattag gcgtgacaca tgacaaatca cagactcaag 1680  
 caagataaag caaaatgatg tgtacataaa actccagagc tatatgtcat attgcaaaaa 1740  
 20 gaggagagct tataagacaa ggcattgactc acaaaaattc atttgccttt cgtgtcaaaa 1800  
 agaggagggc ttacattat ccatgtcata ttgcaaaaga aagagagaaa gaacaacaca 1860  
 atgctgcgtc aattatacat atctgtatgt ccatcattat tcattcacct ttcgtgtacc 1920  
 25 acatttcata tatcatgagt cacttcattg ctggacatta acaaactcta tottaacatt 1980  
 tagatgcaag agcctttatc tcaactataa tgacagatga tttctcattg tttctcaca 2040  
 aaagcattca gtccattagt cctacaacaa c 2071  
 30

&lt;210&gt; 68

&lt;211&gt; 79

&lt;212&gt; PRT

35 &lt;213&gt; Zea mays

&lt;220&gt;

&lt;221&gt; SIGNAL

40 &lt;222&gt; (1)..(79)

&lt;223&gt; Maize waxy signal sequence.

&lt;400&gt; 68

45 Met Leu Ala Ala Leu Ala Thr Ser Gln Leu Val Ala Thr Arg Ala Gly  
 1 5 10 15

50 Leu Gly Val Pro Asp Ala Ser Thr Phe Arg Arg Gly Ala Ala Gln Gly  
 20 25 30

55 Leu Arg Gly Ala Arg Ala Ser Ala Ala Ala Asp Thr Leu Ser Met Arg  
 35 40 45

Thr Ser Ala Arg Ala Ala Pro Arg His Gln His Gln Gln Ala Arg Arg

[illegible]

tgc ggc gcc tgc tgg gcc ttc gcc gcc atc gcc acc gtg gag tcc atc 480  
 Cys Gly Ala Cys Trp Ala Phe Ala Ala Ile Ala Thr Val Glu Ser Ile  
 145 150 155 160

5

tac aag atc aag aag ggc atc ctc gag ccg ctc tcc gag cag cag gtg 528  
 Tyr Lys Ile Lys Lys Gly Ile Leu Glu Pro Leu Ser Glu Gln Gln Val  
 165 170 175

10

ctc gac tgc gcc aag ggc tac ggc tgc aag ggc gcc tgg gag ttc cgc 576  
 Leu Asp Cys Ala Lys Gly Tyr Gly Cys Lys Gly Gly Trp Glu Phe Arg  
 180 185 190

15

gcc ttc gag ttc atc atc tcc aac aag ggc gtg gcc tcc ggc gcc atc 624  
 Ala Phe Glu Phe Ile Ile Ser Asn Lys Gly Val Ala Ser Gly Ala Ile  
 195 200 205

20

tac ccg tac aag gcc gcc aag ggc acc tgc aag acc gac ggc gtg ccg 672  
 Tyr Pro Tyr Lys Ala Ala Lys Gly Thr Cys Lys Thr Asp Gly Val Pro  
 210 215 220

25

aac tcc gcc tac atc acc ggc tac gcc cgc gtg ccg cgc aac aac gag 720  
 Asn Ser Ala Tyr Ile Thr Gly Tyr Ala Arg Val Pro Arg Asn Asn Glu  
 225 230 235 240

tcc tcc atg atg tac gcc gtg tcc aag cag ccg atc acc gtg gcc gtg 768  
 Ser Ser Met Met Tyr Ala Val Ser Lys Gln Pro Ile Thr Val Ala Val  
 245 250 255

30

gac gcc aac gcc aac ttc cag tac tac aag tcc ggc gtg ttc aac ggc 816  
 Asp Ala Asn Ala Asn Phe Gln Tyr Tyr Lys Ser Gly Val Phe Asn Gly  
 260 265 270

35

ccg tgc ggc acc tcc ctc aac cac gcc gtg acc gcc atc ggc tac ggc 864  
 Pro Cys Gly Thr Ser Leu Asn His Ala Val Thr Ala Ile Gly Tyr Gly  
 275 280 285

40

cag gac tcc atc atc tac ccg aag aag tgg ggc gcc aag tgg ggc gag 912  
 Gln Asp Ser Ile Ile Tyr Pro Lys Lys Trp Gly Ala Lys Trp Gly Glu  
 290 295 300

45

gcc ggc tac atc cgc atg gcc cgc gac gtg tcc tcc tcc tcc ggc atc 960  
 Ala Gly Tyr Ile Arg Met Ala Arg Asp Val Ser Ser Ser Ser Gly Ile  
 305 310 315 320

tgc ggc atc gcc atc gac ccg ctc tac ccg acc ctc gag gag tag 1005  
 Cys Gly Ile Ala Ile Asp Pro Leu Tyr Pro Thr Leu Glu Glu  
 325 330

50

<210> 70  
 <211> 334  
 <212> PRT  
 <213> Artificial Sequence

55

<220>  
 <223> Synthetic Construct

```

<400> 70

Met Ala Trp Lys Val Gln Val Val Phe Leu Phe Leu Phe Leu Cys Val
5 1 5 10 15

Met Trp Ala Ser Pro Ser Ala Ala Ser Ala Asp Glu Pro Ser Asp Pro
10 20 25 30

Met Met Lys Arg Phe Glu Glu Trp Met Val Glu Tyr Gly Arg Val Tyr
35 40 45

Lys Asp Asn Asp Glu Lys Met Arg Arg Phe Gln Ile Phe Lys Asn Asn
15 50 55 60

Val Asn His Ile Glu Thr Phe Asn Ser Arg Asn Glu Asn Ser Tyr Thr
20 65 70 75 80

Leu Gly Ile Asn Gln Phe Thr Asp Met Thr Asn Asn Glu Phe Ile Ala
25 85 90 95

Gln Tyr Thr Gly Gly Ile Ser Arg Pro Leu Asn Ile Glu Arg Glu Pro
30 100 105 110

Val Val Ser Phe Asp Asp Val Asp Ile Ser Ala Val Pro Gln Ser Ile
115 120 125

Asp Trp Arg Asp Tyr Gly Ala Val Thr Ser Val Lys Asn Gln Asn Pro
35 130 135 140

Cys Gly Ala Cys Trp Ala Phe Ala Ala Ile Ala Thr Val Glu Ser Ile
40 145 150 155 160

Tyr Lys Ile Lys Lys Gly Ile Leu Glu Pro Leu Ser Glu Gln Gln Val
45 165 170 175

Leu Asp Cys Ala Lys Gly Tyr Gly Cys Lys Gly Gly Trp Glu Phe Arg
180 185 190

Ala Phe Glu Phe Ile Ile Ser Asn Lys Gly Val Ala Ser Gly Ala Ile
195 200 205

Tyr Pro Tyr Lys Ala Ala Lys Gly Thr Cys Lys Thr Asp Gly Val Pro
55 210 215 220

```

5 Asn Ser Ala Tyr Ile Thr Gly Tyr Ala Arg Val Pro Arg Asn Asn Glu  
 225 230 235 240  
 Ser Ser Met Met Tyr Ala Val Ser Lys Gln Pro Ile Thr Val Ala Val  
 245 250 255  
 10 Asp Ala Asn Ala Asn Phe Gln Tyr Tyr Lys Ser Gly Val Phe Asn Gly  
 260 265 270  
 15 Pro Cys Gly Thr Ser Leu Asn His Ala Val Thr Ala Ile Gly Tyr Gly  
 275 280 285  
 20 Gln Asp Ser Ile Ile Tyr Pro Lys Lys Trp Gly Ala Lys Trp Gly Glu  
 290 295 300  
 Ala Gly Tyr Ile Arg Met Ala Arg Asp Val Ser Ser Ser Ser Gly Ile  
 305 310 315 320  
 25 Cys Gly Ile Ala Ile Asp Pro Leu Tyr Pro Thr Leu Glu Glu  
 325 330  
 30 <210> 71  
 <211> 78  
 <212> DNA  
 <213> Artificial Sequence  
 35 <220>  
 <223> Bromealin signal sequence  
 <400> 71  
 40 atggcctgga aggtgcaggt ggtgttcttc ttcctcttcc tctgcgtgat gtgggcctcc 60  
 ccgtccgccc cctccgcc 78  
 45 <210> 72  
 <211> 26  
 <212> PRT  
 <213> Artificial Sequence  
 50 <220>  
 <223> Bromealin signal peptide  
 <400> 72  
 55 Met Ala Trp Lys Val Gln Val Val Phe Leu Phe Leu Phe Leu Cys Val  
 1 5 10 15

Met Trp Ala Ser Pro Ser Ala Ala Ser Ala  
                   20                                  25

5  
     <210> 73  
     <211> 1050  
     <212> DNA  
     <213> Artificial Sequence

10  
     <220>  
     <223> pSYN11000

15  
     <400> 73  
     atggcctgga aggtgcaggt ggtgttcctc ttctctcttc tctgcgtgat gtgggcctcc 60  
     ccgtccgccg cctccgcgga cgagccgtcc gacccgatga tgaagcgctt cgaggagtgg 120  
     atggtggagt acggccgcgt gtacaaggac aacgacgaga agatgcgccg cttccagatc 180  
     20 ttcaagaaca acgtgaacca catcgagacc ttcaactccc gcaacgagaa ctctacacc 240  
     ctcggcatca accagttcac cgacatgacc aacaacgagt tcatcgccca gtacaccggc 300  
     25 ggcatctccc gcccgctcaa catcgagcgc gagccggtgg tgccttcga cgacgtggac 360  
     atctccgccg tgccgcagtc catcgactgg cgcgactacg gcgccgtgac ctccgtgaag 420  
     aaccagaacc cgtgcggcgc ctgctggggc ttgcgcgcca tcgccaccgt ggagtccatc 480  
     30 tacaagatca agaagggcat cctcgagccg ctctccgagc agcaggtgct cgactgcgcc 540  
     aagggctacg gctgcaaggg cggctgggag ttccgcgcct tcgagttcat catctccaac 600  
     35 aagggcgtag cctccggcgc catctacccg tacaaggccg ccaagggcac ctgcaagacc 660  
     gacggcgtgc cgaactccgc ctacatcacc ggctacgcc gcgtgccgcg caacaacgag 720  
     tcctccatga tgtacgccgt gtccaagcag ccgatcaccg tggccgtgga cgccaacgcc 780  
     40 aacttccagt actacaagtc cggcgtgttc aacggcccgt gcggcacctc cctcaaccac 840  
     gccgtgaccg ccatcggtta cggccaggac tccatcatct acccgaagaa gtggggcgcc 900  
     45 aagtggggcg aggccggcta catccgcatg gcccgcgacg tgcctcctc ctccggcatc 960  
     tgcggcatcg ccatcgacc gctctacccg accctcgagg aggtgttcgc cgaggccatc 1020  
     gccgccaact ccaccctcgt ggccgagtag 1050

50  
     <210> 74  
     <211> 1067  
     <212> DNA  
     55 <213> Artificial Sequence

    <220>



&lt;223&gt; pSYN11589

&lt;400&gt; 74

5 tggcctggaa ggtgcagggtg gtgttcctct tcctcttcct ctgcgtgatg tgggcctccc 60

cgctccgccg ctccgcctcc tcctctctct tcgccgactc caacccgatc cgcccgggtga 120

ccgaccgcgc cgctccacc gacgagccgt ccgaccgat gatgaagcgc ttcgaggagt 180

10 ggatggtgga gtacggccgc gtgtacaagg acaacgacga gaagatgcgc cgcttcacaga 240

tcttcaagaa caacgtgaac cacatcgaga ccttcaactc ccgcaacgag aactcctaca 300

ccctcggcat caaccagttc accgacatga ccaacaacga gttcatcgcc cagtacaccg 360

15 gcggcatctc ccgcccgctc aacatcgagc gcgagccggt ggtgtccttc gacgacgtgg 420

acatctccgc cgtgccgcag tccatcgact ggcgcgacta ccgcccgtg acctccgtga 480

20 agaaccagaa cccgtgcggc gcctgctggg ccttcgccgc catcgccacc gtggagtcca 540

tctacaagat caagaagggc atcctcgagc cgctctccga gcagcagggtg ctcgactgcg 600

ccaagggcta cggctgcaag ggcggtggg agttccgcgc cttcgagttc atcatctcca 660

25 acaagggcgt ggctccggc gccatctacc cgtacaaggc cgccaagggc acctgcaaga 720

ccgacggcgt gccgaactcc gcctacatca ccggctacgc ccgctgccc cgcaacaacg 780

30 agtcctccat gatgtacgcc gtgtccaagc agccgatcac cgtggccgtg gacgccaacg 840

ccaacttcca gtactacaag tcggcggtgt tcaacggccc gtgcggcacc tccctcaacc 900

acgccgtgac cgccatcggc tacggccagg actccatcat ctaccgaag aagtggggcg 960

35 ccaagtgggg cgaggccggc tacatccgca tggcccgcga cgtgtcctcc tcctccggca 1020

tctgcccgat cgccatcgac ccgctctacc cgaccctcga ggagtag 1067

40

&lt;210&gt; 75

&lt;211&gt; 1023

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

45

&lt;220&gt;

&lt;223&gt; pSYN11587 Sequence

&lt;400&gt; 75

50 atggcctgga aggtgcagggt ggtgttcctc tcctcttcc tctgcgtgat gtgggcctcc 60

ccgtccgccg cctccgcgga cgagccgtcc gaccgatga tgaagcgctt cgaggagtgg 120

atggtggagt acggccgcgt gtacaaggac aacgacgaga agatgcgccg cttccagatc 180

55 ttcaagaaca acgtgaacca catcgagacc ttcaactccc gcaacgagaa ctccacacc 240

	ctcggcatca accagttcac cgacatgacc aacaacgagt tcatcgccca gtacaccggc	300
	ggcatctccc gcccgctcaa catcgagcgc gagccggtgg tgccttcga cgacgtggac	360
5	atctccgccg tgccgcagtc catcgactgg cgcgactacg gcgccgtgac ctccgtgaag	420
	aaccagaacc cgtgcggcgc ctgctgggccc ttccgcgcca tcgccaccgt ggagtccatc	480
	tacaagatca agaagggcat cctcgagccg ctctccgagc agcaggtgct cgactgcgcc	540
10	aagggctacg gctgcaaggg cggtcgggag ttccgcgcct tcgagttcat catctccaac	600
	aagggcgtgg cctccggcgc catctacccg tacaaggccg ccaagggcac ctgcaagacc	660
15	gacggcgtgc cgaactccgc ctacatcacc ggctacgccc gcgtgccgcg caacaacgag	720
	tcctccatga tgtacgccgt gtccaagcag ccgatcaccg tggccgtgga cgccaacgcc	780
	aacttccagt actacaagtc cggcgtgttc aacggcccgt gcggcacctc cctcaaccac	840
20	gccgtgaccg ccatcggcta cggccaggac tccatcatct acccgaagaa gtggggcgcc	900
	aagtggggcg aggccggcta catccgcatt gcccgcgacg tgcctcctc ctccggcatc	960
25	tgcggcatcg ccatcgaccc gctctacccg accctcgagg agtccgagaa ggacgagctg	1020
	tag	1023
30	<210> 76	
	<211> 990	
	<212> DNA	
	<213> Artificial Sequence	
35	<220>	
	<223> pSYN12169 Sequence	
	<400> 76	
	atgagggtgt tgctcgttgc cctcgctctc ctggctctcg ctgcgagcgc cacctccatg	60
40	gcggacgagc cgtccgaccc gatgatgaag cgcttcgagg agtggatggt ggagtacggc	120
	cgcgtgtaca aggacaacga cgagaagatg cgccgcttcc agatcttcaa gaacaacgtg	180
45	aaccacatcg agaccttcaa ctcccgaac gagaactcct acaccctcgg catcaaccag	240
	ttcaccgaca tgaccaacaa cgagttcatc gccagtaga ccggcgggcat ctcccgccc	300
	ctcaacatcg agcgcgagcc ggtgggtgtcc ttcgacgacg tggacatctc cgccgtgccg	360
50	cagtccatcg actggcgga ctacggcgcc gtgacctcg tgaagaacca gaaccctgac	420
	ggcgctgct gggccttcgc cgccatcgcc accgtggagt ccatctacaa gatcaagaag	480
55	ggcatcctcg agccgctctc cgagcagcag gtgctcgact gcgccaaggg ctacggctgc	540
	aagggcggct gggagttccg cgccttcgag ttcatcatct ccaacaaggg cgtggcctcc	600

ggcgccatct acccggtacaa ggccgccaag ggcacctgca agaccgacgg cgtgccgaac 660  
 tccgcctaca tcaccggcta cgcgcgctg cgcgcgaaca acgagtcctc catgatgtac 720  
 5 gccgtgtcca agcagccgat caccgtggcc gtggacgcca acgccaactt ccagtactac 780  
 aagtccggcg tgttcaacgg ccggtgcggc acctccctca accacgccgt gaccgccatc 840  
 10 ggctacggcc aggactccat catctaccg aagaagtggg gcgccaagtg gggcgaggcc 900  
 ggctacatcc gcatggcccg cgacgtgtcc tcctcctccg gcatctgcgg catcgccatc 960  
 gaccgcctct acccgaccct cgaggagtag 990  
 15  
 <210> 77  
 <211> 1170  
 <212> DNA  
 20 <213> Artificial Sequence  
 <220>  
 <223> pSYN12575 Sequence  
 25 <400> 77  
 atgctggcgg ctctggccac gtgcgagctc gtgcgaacgc gcgccggcct gggcgctccc 60  
 gacgcgtcca cgttccgcgg cggcgccgcg cagggcctga ggggggcccg ggcgtcggcg 120  
 30 gcggcggaca cgctcagcat gcggaccagc gcgcgcgcgg cggccaggca ccagcaccag 180  
 caggcgcgcc gcggggccag gttcccgtcg ctcgctgtgt gcgccagcgc cggcgccatg 240  
 gcggacgagc cgtccgaccc gatgatgaag cgcttcgagg agtggatggg ggagtacggc 300  
 35 cgcgtgtaca aggacaacga cgagaagatg cgcgcgttcc agatcttcaa gaacaacgtg 360  
 aaccacatcg agaccttcaa ctcccgcaac gagaactcct acaccctcgg catcaaccag 420  
 40 ttacccgaca tgaccaaaaa cgagttcatc gccagctaca ccggcggcat ctcccgcccg 480  
 ctcaacatcg agcgcgagcc ggtgggtgtcc ttcgacgacg tggacatctc cgccgtgccg 540  
 cagtccatcg actggcgcgga ctacggcgcc gtgacctcgg tgaagaacca gaaccgtgc 600  
 45 ggcgcctgct gggccttcgc cgccatcgcc accgtggagt ccatctacaa gatcaagaag 660  
 ggcacctcgg agccgctctc cgagcagcag gtgctcgact gcgccaaggg ctacggctgc 720  
 50 aaggcgggct gggagttccg cgcttcgag ttcacatct ccaacaaggg cgtggcctcc 780  
 ggcgccatct acccggtacaa ggccgccaag ggcacctgca agaccgacgg cgtgccgaac 840  
 tccgcctaca tcaccggcta cgcgcgctg cgcgcgaaca acgagtcctc catgatgtac 900  
 55 gccgtgtcca agcagccgat caccgtggcc gtggacgcca acgccaactt ccagtactac 960

aagtccggcg tgttcaacgg cccgtgcggc acctccctca accacgccgt gaccgccatc 1020  
 ggctacggcc aggactccat catctacccg aagaagtggg gcgccaagtg gggcgaggcc 1080  
 5 ggctacatcc gcatggcccc cgacgtgtcc tctcctccg gcctctgcgg catcgccatc 1140  
 gaccgcctct acccgaccct cgaggagtag 1170

10 <210> 78  
 <211> 1068  
 <212> DNA  
 <213> Artificial Sequence

15 <220>  
 <223> pSM270 Sequence

<400> 78  
 atggcctgga aggtgcaggt ggtgttcctc ttcctcttcc tctgcgtgat gtgggcctcc 60  
 20 ccgtccgcgg cctccgcctc ctctcctcc ttcgcgact ccaaccgat ccgcccggcg 120  
 accgaccgag ccgcctccac cgacgagccg tccgaccga tgatgaagcg ctctgaggag 180  
 25 tggatggtgg agtacggccg cgtgtacaag gacaacgacg agaagatgcg ccgcttccag 240  
 atcttcaaga acaacgtgaa ccacatcgag accttcaact cccgcaacga gaactcctac 300  
 accctcggca tcaaccagtt caccgacatg accaacaacg agttcatcgc ccagtacacc 360  
 30 ggcggcatct cccgcccgtt caacatcgag cgcgagccgg tgggtgtcctt cgacgacgtg 420  
 gacatctccg ccgtgccgca gtccatcgac tggcgcgact acggcgccgt gacctccgtg 480  
 35 aagaaccaga acccgtgcgg cgcttgcctg gccttcgccc ccatcgccac cgtggagtc 540  
 atctacaaga tcaagaaggg catcctcgag ccgctctccg agcagcaggt gctcgactgc 600  
 gccaaagggt acggctgcaa gggcggtcgg gagttccgag ccttcgagtt catcatctcc 660  
 40 aacaagggag tggcctccgg cgccatctac ccgtacaagg ccgccaaggg cacctgcaag 720  
 accgacggcg tgccgaactc cgctacatc accggctacg cccgcgtgcc gcgcaacaac 780  
 45 gagtctccca tgatgtacgc cgtgtccaag cagccgatca ccgtggccgt ggacgccaac 840  
 gccaaacttc agtactacaa gtccggcggt ttcaacggcc cgtgcggcac ctccctcaac 900  
 cagcccgtag ccgccatcgg ctacggccag gactccatca tctacccgaa gaagtggggc 960  
 50 gccaaagtgg gcgaggccgg ctacatccgc atggcccgcg acgtgtcctc ctctccggc 1020  
 atctgcggca tcgcatcga cccgctctac ccgacctcg aggagtag 1068

55 <210> 79  
 <211> 1497

<212> DNA  
<213> *Trichoderma reesei*

5 <220>  
<221> CDS  
<222> (1)..(1497)  
<223> *Trichoderma reesei* cellobiohydrolase I

10 <400> 79  
atg cag tcg gcg tgt act ctc caa tcg gag act cac ccg cct ctg aca 48  
Met Gln Ser Ala Cys Thr Leu Gln Ser Glu Thr His Pro Pro Leu Thr  
1 5 10 15

15 tgg cag aaa tgc tcg tct ggt ggc acg tgc act caa cag aca ggc tcc 96  
Trp Gln Lys Cys Ser Ser Gly Gly Thr Cys Thr Gln Gln Thr Gly Ser  
20 25 30

20 gtg gtc atc gac gcc aac tgg cgc tgg act cac gct acg aac agc agc 144  
Val Val Ile Asp Ala Asn Trp Arg Trp Thr His Ala Thr Asn Ser Ser  
35 40 45

25 acg aac tgc tac gat ggc aac act tgg agc tcg acc cta tgt cct gac 192  
Thr Asn Cys Tyr Asp Gly Asn Thr Trp Ser Ser Thr Leu Cys Pro Asp  
50 55 60

30 aac gag acc tgc gcg aag aac tgc tgt ctg gac ggt gcc gcc tac gcg 240  
Asn Glu Thr Cys Ala Lys Asn Cys Cys Leu Asp Gly Ala Ala Tyr Ala  
65 70 75 80

35 tcc acg tac gga gtt acc acg agc ggt aac agc ctc tcc att ggc ttt 288  
Ser Thr Tyr Gly Val Thr Thr Ser Gly Asn Ser Leu Ser Ile Gly Phe  
85 90 95

40 gtc acc cag tct gcg cag aag aac gtt ggc gct cgc ctt tac ctt atg 336  
Val Thr Gln Ser Ala Gln Lys Asn Val Gly Ala Arg Leu Tyr Leu Met  
100 105 110

45 gcg agc gac acg acc tac cag gaa ttc acc ctg ctt ggc aac gag ttc 384  
Ala Ser Asp Thr Thr Tyr Gln Glu Phe Thr Leu Leu Gly Asn Glu Phe  
115 120 125

50 tct ttc gat gtt gat gtt tcg cag ctg ccg tgc ggc ttg aac gga gct 432  
Ser Phe Asp Val Asp Val Ser Gln Leu Pro Cys Gly Leu Asn Gly Ala  
130 135 140

55 ctc tac ttc gtg tcc atg gac gcg gat ggt ggc gtg agc aag tat ccc 480  
Leu Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Val Ser Lys Tyr Pro  
145 150 155 160

60 acc aac acc gct ggc gcc aag tac ggc acg ggg tac tgt gac agc cag 528  
Thr Asn Thr Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln  
165 170 175

65 tgt ccc cgc gat ctg aag ttc atc aat ggc cag gcc aac gtt gag ggc 576  
Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val Glu Gly  
180 185 190

	tgg gag ccg tca tcc aac aac gcg aac acg ggc att gga gga cac gga	624
	Trp Glu Pro Ser Ser Asn Asn Ala Asn Thr Gly Ile Gly Gly His Gly	
	195 200 205	
5	agc tgc tgc tct gag atg gat atc tgg gag gcc aac tcc atc tcc gag	672
	Ser Cys Cys Ser Glu Met Asp Ile Trp Glu Ala Asn Ser Ile Ser Glu	
	210 215 220	
10	gct ctt acc ccc cac cct tgc acg act gtc ggc cag gag atc tgc gag	720
	Ala Leu Thr Pro His Pro Cys Thr Thr Val Gly Gln Glu Ile Cys Glu	
	225 230 235 240	
15	ggg gat ggg tgc ggc gga act tac tcc gat aac aga tat ggc ggc act	768
	Gly Asp Gly Cys Gly Gly Thr Tyr Ser Asp Asn Arg Tyr Gly Gly Thr	
	245 250 255	
20	tgc gat ccc gat ggc tgc gac tgg aac cca tac cgc ctg ggc aac acc	816
	Cys Asp Pro Asp Gly Cys Asp Trp Asn Pro Tyr Arg Leu Gly Asn Thr	
	260 265 270	
	agc ttc tac ggc cct ggc tct agc ttt acc ctc gat acc acc aag aaa	864
	Ser Phe Tyr Gly Pro Gly Ser Ser Phe Thr Leu Asp Thr Thr Lys Lys	
	275 280 285	
25	ttg acc gtt gtc acc cag ttc gag acg tcg ggt gcc atc aac cga tac	912
	Leu Thr Val Val Thr Gln Phe Glu Thr Ser Gly Ala Ile Asn Arg Tyr	
	290 295 300	
30	tat gtc cag aat ggc gtc act ttc cag cag ccc aac gcc gag ctt ggt	960
	Tyr Val Gln Asn Gly Val Thr Phe Gln Gln Pro Asn Ala Glu Leu Gly	
	305 310 315 320	
35	agt tac tct ggc aac gag ctc aac gat gat tac tgc aca gct gag gag	1008
	Ser Tyr Ser Gly Asn Glu Leu Asn Asp Asp Tyr Cys Thr Ala Glu Glu	
	325 330 335	
40	gca gaa ttc ggc gga tcc tct ttc tca gac aag ggc ggc ctg act cag	1056
	Ala Glu Phe Gly Gly Ser Ser Phe Ser Asp Lys Gly Gly Leu Thr Gln	
	340 345 350	
	ttc aag aag gct acc tct ggc ggc atg gtt ctg gtc atg agt ctg tgg	1104
	Phe Lys Lys Ala Thr Ser Gly Gly Met Val Leu Val Met Ser Leu Trp	
	355 360 365	
45	gat gat tac tac gcc aac atg ctg tgg ctg gac tcc acc tac ccg aca	1152
	Asp Asp Tyr Tyr Ala Asn Met Leu Trp Leu Asp Ser Thr Tyr Pro Thr	
	370 375 380	
50	aac gag acc tcc tcc aca ccc ggt gcc gtg cgc gga agc tgc tcc acc	1200
	Asn Glu Thr Ser Ser Thr Pro Gly Ala Val Arg Gly Ser Cys Ser Thr	
	385 390 395 400	
55	agc tcc ggt gtc cct gct cag gtc gaa tct cag tct ccc aac gcc aag	1248
	Ser Ser Gly Val Pro Ala Gln Val Glu Ser Gln Ser Pro Asn Ala Lys	
	405 410 415	

gtc acc ttc tcc aac atc aag ttc gga ccc att ggc agc acc ggc aac 1296  
 Val Thr Phe Ser Asn Ile Lys Phe Gly Pro Ile Gly Ser Thr Gly Asn  
 420 425 430

5 cct agc ggc ggc aac cct ccc ggc gga aac ccg cct ggc acc acc acc 1344  
 Pro Ser Gly Gly Asn Pro Pro Gly Gly Asn Pro Pro Gly Thr Thr Thr  
 435 440 445

10 acc cgc cgc cca gcc act acc act gga agc tct ccc gga cct acc cag 1392  
 Thr Arg Arg Pro Ala Thr Thr Thr Gly Ser Ser Pro Gly Pro Thr Gln  
 450 455 460

15 tct cac tac ggc cag tgc ggc ggt att ggc tac agc ggc ccc acg gtc 1440  
 Ser His Tyr Gly Gln Cys Gly Gly Ile Gly Tyr Ser Gly Pro Thr Val  
 465 470 475 480

tgc gcc agc ggc aca act tgc cag gtc ctg aac cct tac tac tct cag 1488  
 Cys Ala Ser Gly Thr Thr Cys Gln Val Leu Asn Pro Tyr Tyr Ser Gln  
 485 490 495

20 tgc ctg taa 1497  
 Cys Leu

25 <210> 80  
 <211> 498  
 <212> PRT  
 <213> Trichoderma reesei

30 <400> 80

Met Gln Ser Ala Cys Thr Leu Gln Ser Glu Thr His Pro Pro Leu Thr  
 1 5 10 15

35 Trp Gln Lys Cys Ser Ser Gly Gly Thr Cys Thr Gln Gln Thr Gly Ser  
 20 25 30

40 Val Val Ile Asp Ala Asn Trp Arg Trp Thr His Ala Thr Asn Ser Ser  
 35 40 45

45 Thr Asn Cys Tyr Asp Gly Asn Thr Trp Ser Ser Thr Leu Cys Pro Asp  
 50 55 60

50 Asn Glu Thr Cys Ala Lys Asn Cys Cys Leu Asp Gly Ala Ala Tyr Ala  
 65 70 75 80

Ser Thr Tyr Gly Val Thr Thr Ser Gly Asn Ser Leu Ser Ile Gly Phe  
 85 90 95

55 Val Thr Gln Ser Ala Gln Lys Asn Val Gly Ala Arg Leu Tyr Leu Met

	100	105	110
5	Ala Ser Asp Thr Thr Tyr Gln Glu Phe Thr Leu Leu Gly Asn Glu Phe 115 120 125		
10	Ser Phe Asp Val Asp Val Ser Gln Leu Pro Cys Gly Leu Asn Gly Ala 130 135 140		
15	Leu Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Val Ser Lys Tyr Pro 145 150 155 160		
20	Thr Asn Thr Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln 165 170 175		
25	Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val Glu Gly 180 185 190		
30	Trp Glu Pro Ser Ser Asn Asn Ala Asn Thr Gly Ile Gly Gly His Gly 195 200 205		
35	Ser Cys Cys Ser Glu Met Asp Ile Trp Glu Ala Asn Ser Ile Ser Glu 210 215 220		
40	Ala Leu Thr Pro His Pro Cys Thr Thr Val Gly Gln Glu Ile Cys Glu 225 230 235 240		
45	Gly Asp Gly Cys Gly Gly Thr Tyr Ser Asp Asn Arg Tyr Gly Gly Thr 245 250 255		
50	Cys Asp Pro Asp Gly Cys Asp Trp Asn Pro Tyr Arg Leu Gly Asn Thr 260 265 270		
55	Ser Phe Tyr Gly Pro Gly Ser Ser Phe Thr Leu Asp Thr Thr Lys Lys 275 280 285		
60	Leu Thr Val Val Thr Gln Phe Glu Thr Ser Gly Ala Ile Asn Arg Tyr 290 295 300		
65	Tyr Val Gln Asn Gly Val Thr Phe Gln Gln Pro Asn Ala Glu Leu Gly 305 310 315 320		
70	Ser Tyr Ser Gly Asn Glu Leu Asn Asp Asp Tyr Cys Thr Ala Glu Glu 325 330 335		



Ala Glu Phe Gly Gly Ser Ser Phe Ser Asp Lys Gly Gly Leu Thr Gln  
340 345 350

5

Phe Lys Lys Ala Thr Ser Gly Gly Met Val Leu Val Met Ser Leu Trp  
355 360 365

10

Asp Asp Tyr Tyr Ala Asn Met Leu Trp Leu Asp Ser Thr Tyr Pro Thr  
370 375 380

15

Asn Glu Thr Ser Ser Thr Pro Gly Ala Val Arg Gly Ser Cys Ser Thr  
385 390 395 400

20

Ser Ser Gly Val Pro Ala Gln Val Glu Ser Gln Ser Pro Asn Ala Lys  
405 410 415

25

Val Thr Phe Ser Asn Ile Lys Phe Gly Pro Ile Gly Ser Thr Gly Asn  
420 425 430

30

Pro Ser Gly Gly Asn Pro Pro Gly Gly Asn Pro Pro Gly Thr Thr Thr  
435 440 445

35

Thr Arg Arg Pro Ala Thr Thr Thr Gly Ser Ser Pro Gly Pro Thr Gln  
450 455 460

40

Ser His Tyr Gly Gln Cys Gly Gly Ile Gly Tyr Ser Gly Pro Thr Val  
465 470 475 480

45

Cys Ala Ser Gly Thr Thr Cys Gln Val Leu Asn Pro Tyr Tyr Ser Gln  
485 490 495

Cys Leu

50

<210> 81  
<211> 1365  
<212> DNA  
<213> Trichoderma reesei

55

<220>  
<221> CDS  
<222> (1)..(1365)  
<223> trichoderma reesei cellobiohydrolase II

<400> 81																
	atg	gtg	cct	cta	gag	gag	cgg	caa	gct	tgc	tca	agc	gtc	tgg	ggc	caa
	Met	Val	Pro	Leu	Glu	Glu	Arg	Gln	Ala	Cys	Ser	Ser	Val	Trp	Gly	Gln
	1				5				10					15		
5	tgt	ggg	ggc	cag	aat	tgg	tcg	ggg	ccg	act	tgc	tgt	gct	tcc	gga	agc
	Cys	Gly	Gly	Gln	Asn	Trp	Ser	Gly	Pro	Thr	Cys	Cys	Ala	Ser	Gly	Ser
				20					25					30		
10	aca	tgc	gtc	tac	tcc	aac	gac	tat	tac	tcc	cag	tgt	ctt	ccc	ggc	gct
	Thr	Cys	Val	Tyr	Ser	Asn	Asp	Tyr	Tyr	Ser	Gln	Cys	Leu	Pro	Gly	Ala
				35				40						45		
15	gca	agc	tca	agc	tcg	tcc	acg	cgc	gcc	gcg	tcg	acg	act	tca	cga	gta
	Ala	Ser	Ser	Ser	Ser	Ser	Thr	Arg	Ala	Ala	Ser	Thr	Thr	Ser	Arg	Val
		50					55					60				
20	tcc	ccc	aca	aca	tcc	cgg	tcg	agc	tcc	gcg	acg	cct	cca	cct	ggg	tct
	Ser	Pro	Thr	Thr	Ser	Arg	Ser	Ser	Ser	Ala	Thr	Pro	Pro	Pro	Gly	Ser
	65					70					75				80	
25	acc	act	acc	aga	gta	cct	cca	gtc	gga	tcg	gga	acc	gct	acg	tat	tca
	Thr	Thr	Thr	Arg	Val	Pro	Pro	Val	Gly	Ser	Gly	Thr	Ala	Thr	Tyr	Ser
				85					90						95	
30	ggc	aac	cct	ttt	gtt	ggg	gtc	act	cct	tgg	gcc	aat	gca	tat	tac	gcc
	Gly	Asn	Pro	Phe	Val	Gly	Val	Thr	Pro	Trp	Ala	Asn	Ala	Tyr	Tyr	Ala
				100					105					110		
35	tct	gaa	gtt	agc	agc	ctc	gct	att	cct	agc	ttg	act	gga	gcc	atg	gcc
	Ser	Glu	Val	Ser	Ser	Leu	Ala	Ile	Pro	Ser	Leu	Thr	Gly	Ala	Met	Ala
			115					120					125			
40	act	gct	gca	gca	gct	gtc	gca	aag	gtt	ccc	tct	ttt	atg	tgg	cta	gat
	Thr	Ala	Ala	Ala	Ala	Val	Ala	Lys	Val	Pro	Ser	Phe	Met	Trp	Leu	Asp
		130					135					140				
45	act	ctt	gac	aag	acc	cct	ctc	atg	gag	caa	acc	ttg	gcc	gac	atc	cgc
	Thr	Leu	Asp	Lys	Thr	Pro	Leu	Met	Glu	Gln	Thr	Leu	Ala	Asp	Ile	Arg
	145					150					155				160	
50	acc	gcc	aac	aag	aat	ggc	ggg	aac	tat	gcc	gga	cag	ttt	gtg	gtg	tat
	Thr	Ala	Asn	Lys	Asn	Gly	Gly	Asn	Tyr	Ala	Gly	Gln	Phe	Val	Val	Tyr
					165					170				175		
55	gac	ttg	ccg	gat	cgc	gat	tgc	gct	gcc	ctt	gcc	tcg	aat	ggc	gaa	tac
	Asp	Leu	Pro	Asp	Arg	Asp	Cys	Ala	Ala	Leu	Ala	Ser	Asn	Gly	Glu	Tyr
				180				185						190		
60	tct	att	gcc	gat	ggg	ggc	gtc	gcc	aaa	tat	aag	aac	tat	atc	gac	acc
	Ser	Ile	Ala	Asp	Gly	Gly	Val	Ala	Lys	Tyr	Lys	Asn	Tyr	Ile	Asp	Thr
			195					200					205			
65	att	cgt	caa	att	gtc	gtg	gaa	tat	tcc	gat	atc	cgg	acc	ctc	ctg	gtt
	Ile	Arg	Gln	Ile	Val	Val	Glu	Tyr	Ser	Asp	Ile	Arg	Thr	Leu	Leu	Val
		210					215					220				

	att gag cct gac tct ctt gcc aac ctg gtg acc aac ctc ggt act cca	720
	Ile Glu Pro Asp Ser Leu Ala Asn Leu Val Thr Asn Leu Gly Thr Pro	
	225 230 235 240	
5	aag tgt gcc aat gct cag tca gcc tac ctt gag tgc atc aac tac gcc	768
	Lys Cys Ala Asn Ala Gln Ser Ala Tyr Leu Glu Cys Ile Asn Tyr Ala	
	245 250 255	
10	gtc aca cag ctg aac ctt cca aat gtt gcg atg tat ttg gac gct ggc	816
	Val Thr Gln Leu Asn Leu Pro Asn Val Ala Met Tyr Leu Asp Ala Gly	
	260 265 270	
15	cat gca gga tgg ctt ggc tgg ccg gca aac caa gac ccg gcc gct cag	864
	His Ala Gly Trp Leu Gly Trp Pro Ala Asn Gln Asp Pro Ala Ala Gln	
	275 280 285	
20	cta ttt gca aat gtt tac aag aat gca tcg tct ccg aga gct ctt cgc	912
	Leu Phe Ala Asn Val Tyr Lys Asn Ala Ser Ser Pro Arg Ala Leu Arg	
	290 295 300	
	gga ttg gca acc aat gtc gcc aac tac aac ggg tgg aac att acc agc	960
	Gly Leu Ala Thr Asn Val Ala Asn Tyr Asn Gly Trp Asn Ile Thr Ser	
	305 310 315 320	
25	ccc cca tcg tac acg caa ggc aac gct gtc tac aac gag aag ctg tac	1008
	Pro Pro Ser Tyr Thr Gln Gly Asn Ala Val Tyr Asn Glu Lys Leu Tyr	
	325 330 335	
30	atc cac gct att gga cct ctt ctt gcc aat cac ggc tgg tcc aac gcc	1056
	Ile His Ala Ile Gly Pro Leu Leu Ala Asn His Gly Trp Ser Asn Ala	
	340 345 350	
35	ttc ttc atc act gat caa ggt cga tcg gga aag cag cct acc gga cag	1104
	Phe Phe Ile Thr Asp Gln Gly Arg Ser Gly Lys Gln Pro Thr Gly Gln	
	355 360 365	
40	caa cag tgg gga gac tgg tgc aat gtg atc ggc acc gga ttt ggt att	1152
	Gln Gln Trp Gly Asp Trp Cys Asn Val Ile Gly Thr Gly Phe Gly Ile	
	370 375 380	
	cgc cca tcc gca aac act ggg gac tcg ttg ctg gat tcg ttt gtc tgg	1200
	Arg Pro Ser Ala Asn Thr Gly Asp Ser Leu Leu Asp Ser Phe Val Trp	
	385 390 395 400	
45	gtc aag cca ggc ggc gag tgt gac ggc acc agc gac agc agt gcg cca	1248
	Val Lys Pro Gly Gly Glu Cys Asp Gly Thr Ser Asp Ser Ser Ala Pro	
	405 410 415	
50	cga ttt gac tcc cac tgt gcg ctc cca gat gcc ttg caa ccg gcg cct	1296
	Arg Phe Asp Ser His Cys Ala Leu Pro Asp Ala Leu Gln Pro Ala Pro	
	420 425 430	
55	caa gct ggt gct tgg ttc caa gcc tac ttt gtg cag ctt ctc aca aac	1344
	Gln Ala Gly Ala Trp Phe Gln Ala Tyr Phe Val Gln Leu Leu Thr Asn	
	435 440 445	
	gca aac cca tcg ttc ctg tag	1365

Ala Asn Pro Ser Phe Leu  
450

5 <210> 82  
 <211> 454  
 <212> PRT  
 <213> Trichoderma reesei

10 <400> 82

Met Val Pro Leu Glu Glu Arg Gln Ala Cys Ser Ser Val Trp Gly Gln  
 1 5 10 15

15 Cys Gly Gly Gln Asn Trp Ser Gly Pro Thr Cys Cys Ala Ser Gly Ser  
 20 25 30

20 Thr Cys Val Tyr Ser Asn Asp Tyr Tyr Ser Gln Cys Leu Pro Gly Ala  
 35 40 45

Ala Ser Ser Ser Ser Ser Thr Arg Ala Ala Ser Thr Thr Ser Arg Val  
 25 50 55 60

Ser Pro Thr Thr Ser Arg Ser Ser Ser Ala Thr Pro Pro Pro Gly Ser  
 65 70 75 80

30 Thr Thr Thr Arg Val Pro Pro Val Gly Ser Gly Thr Ala Thr Tyr Ser  
 85 90 95

35 Gly Asn Pro Phe Val Gly Val Thr Pro Trp Ala Asn Ala Tyr Tyr Ala  
 100 105 110

40 Ser Glu Val Ser Ser Leu Ala Ile Pro Ser Leu Thr Gly Ala Met Ala  
 115 120 125

Thr Ala Ala Ala Ala Val Ala Lys Val Pro Ser Phe Met Trp Leu Asp  
 45 130 135 140

Thr Leu Asp Lys Thr Pro Leu Met Glu Gln Thr Leu Ala Asp Ile Arg  
 145 150 155 160

50 Thr Ala Asn Lys Asn Gly Gly Asn Tyr Ala Gly Gln Phe Val Val Tyr  
 165 170 175

55 Asp Leu Pro Asp Arg Asp Cys Ala Ala Leu Ala Ser Asn Gly Glu Tyr  
 180 185 190

Ser Ile Ala Asp Gly Gly Val Ala Lys Tyr Lys Asn Tyr Ile Asp Thr  
 195 200 205  
 5

Ile Arg Gln Ile Val Val Glu Tyr Ser Asp Ile Arg Thr Leu Leu Val  
 210 215 220

10

Ile Glu Pro Asp Ser Leu Ala Asn Leu Val Thr Asn Leu Gly Thr Pro  
 225 230 235 240

15

Lys Cys Ala Asn Ala Gln Ser Ala Tyr Leu Glu Cys Ile Asn Tyr Ala  
 245 250 255

20

Val Thr Gln Leu Asn Leu Pro Asn Val Ala Met Tyr Leu Asp Ala Gly  
 260 265 270

His Ala Gly Trp Leu Gly Trp Pro Ala Asn Gln Asp Pro Ala Ala Gln  
 275 280 285

25

Leu Phe Ala Asn Val Tyr Lys Asn Ala Ser Ser Pro Arg Ala Leu Arg  
 290 295 300

30

Gly Leu Ala Thr Asn Val Ala Asn Tyr Asn Gly Trp Asn Ile Thr Ser  
 305 310 315 320

35

Pro Pro Ser Tyr Thr Gln Gly Asn Ala Val Tyr Asn Glu Lys Leu Tyr  
 325 330 335

40

Ile His Ala Ile Gly Pro Leu Leu Ala Asn His Gly Trp Ser Asn Ala  
 340 345 350

Phe Phe Ile Thr Asp Gln Gly Arg Ser Gly Lys Gln Pro Thr Gly Gln  
 355 360 365

45

Gln Gln Trp Gly Asp Trp Cys Asn Val Ile Gly Thr Gly Phe Gly Ile  
 370 375 380

50

Arg Pro Ser Ala Asn Thr Gly Asp Ser Leu Leu Asp Ser Phe Val Trp  
 385 390 395 400

55

Val Lys Pro Gly Gly Glu Cys Asp Gly Thr Ser Asp Ser Ser Ala Pro  
 405 410 415

Arg Phe Asp Ser His Cys Ala Leu Pro Asp Ala Leu Gln Pro Ala Pro  
 420 425 430

5  
 Gln Ala Gly Ala Trp Phe Gln Ala Tyr Phe Val Gln Leu Leu Thr Asn  
 435 440 445

10 Ala Asn Pro Ser Phe Leu  
 450

15 <210> 83  
 <211> 1317  
 <212> DNA  
 <213> *Trichoderma reesei*

20 <220>  
 <221> CDS  
 <222> (1)..(1317)  
 <223> *Trichoderma reesei* endoglucanase I

25 <400> 83  
 atg cag caa ccg gga acc agc acc ccc gag gtc cat ccc aag ttg aca 48  
 Met Gln Gln Pro Gly Thr Ser Thr Pro Glu Val His Pro Lys Leu Thr  
 1 5 10 15

30 acc tac aag tgc aca aag tcc ggg ggg tgc gtg gcc cag gac acc tcg 96  
 Thr Tyr Lys Cys Thr Lys Ser Gly Gly Cys Val Ala Gln Asp Thr Ser  
 20 25 30

35 gtg gtc ctt gac tgg aac tac cgc tgg atg cac gac gca aac tac aac 144  
 Val Val Leu Asp Trp Asn Tyr Arg Trp Met His Asp Ala Asn Tyr Asn  
 35 40 45

40 tcg tgc acc gtc aac ggc ggc gtc aac acc acg ctc tgc cct gac gag 192  
 Ser Cys Thr Val Asn Gly Gly Val Asn Thr Thr Leu Cys Pro Asp Glu  
 50 55 60

45 gcg acc tgt ggc aag aac tgc ttc atc gag ggc gtc gac tac gcc gcc 240  
 Ala Thr Cys Gly Lys Asn Cys Phe Ile Glu Gly Val Asp Tyr Ala Ala  
 65 70 75 80

50 tcg ggc gtc acg acc tcg ggc agc agc ctc acc atg aac cag tac atg 288  
 Ser Gly Val Thr Thr Ser Gly Ser Ser Leu Thr Met Asn Gln Tyr Met  
 85 90 95

50 ccc agc agc tct ggc ggc tac agc agc gtc tct cct cgg ctg tat ctc 336  
 Pro Ser Ser Ser Gly Gly Tyr Ser Ser Val Ser Pro Arg Leu Tyr Leu  
 100 105 110

55 ctg gac tct gac ggt gag tac gtg atg ctg aag ctc aac ggc cag gag 384  
 Leu Asp Ser Asp Gly Glu Tyr Val Met Leu Lys Leu Asn Gly Gln Glu  
 115 120 125

	ctg agc ttc gac gtc gac ctc tct gct ctg ccg tgt gga gag aac ggc	432
	Leu Ser Phe Asp Val Asp Leu Ser Ala Leu Pro Cys Gly Glu Asn Gly	
	130 135 140	
5	tcg ctc tac ctg tct cag atg gac gag aac ggg ggc gcc aac cag tat	480
	Ser Leu Tyr Leu Ser Gln Met Asp Glu Asn Gly Gly Ala Asn Gln Tyr	
	145 150 155 160	
10	aac acg gcc ggt gcc aac tac ggg agc ggc tac tgc gat gct cag tgc	528
	Asn Thr Ala Gly Ala Asn Tyr Gly Ser Gly Tyr Cys Asp Ala Gln Cys	
	165 170 175	
15	ccc gtc cag aca tgg agg aac ggc acc ctc aac act agc cac cag ggc	576
	Pro Val Gln Thr Trp Arg Asn Gly Thr Leu Asn Thr Ser His Gln Gly	
	180 185 190	
20	ttc tgc tgc aac gag atg gat atc ctg gag ggc aac tcg agg gcg aat	624
	Phe Cys Cys Asn Glu Met Asp Ile Leu Glu Gly Asn Ser Arg Ala Asn	
	195 200 205	
25	gcc ttg acc cct cac tct tgc acg gcc acg gcc tgc gac tct gcc ggt	672
	Ala Leu Thr Pro His Ser Cys Thr Ala Thr Ala Cys Asp Ser Ala Gly	
	210 215 220	
30	tgc ggc ttc aac ccc tat ggc agc ggc tac aaa agc tac tac ggc ccc	720
	Cys Gly Phe Asn Pro Tyr Gly Ser Gly Tyr Lys Ser Tyr Tyr Gly Pro	
	225 230 235 240	
35	gga gat acc gtt gac acc tcc aag acc ttc acc atc atc acc cag ttc	768
	Gly Asp Thr Val Asp Thr Ser Lys Thr Phe Thr Ile Ile Thr Gln Phe	
	245 250 255	
40	aac acg gac aac ggc tcg ccc tcg ggc aac ctt gtg agc atc acc cgc	816
	Asn Thr Asp Asn Gly Ser Pro Ser Gly Asn Leu Val Ser Ile Thr Arg	
	260 265 270	
45	aag tac cag caa aac ggc gtc gac atc ccc agc gcc cag ccc ggc ggc	864
	Lys Tyr Gln Gln Asn Gly Val Asp Ile Pro Ser Ala Gln Pro Gly Gly	
	275 280 285	
50	gac acc atc tcg tcc tgc ccg tcc gcc tca gcc tac ggc ggc ctc gcc	912
	Asp Thr Ile Ser Ser Cys Pro Ser Ala Ser Ala Tyr Gly Gly Leu Ala	
	290 295 300	
55	acc atg ggc aag gcc ctg agc agc ggc atg gtg ctc gtg ttc agc att	960
	Thr Met Gly Lys Ala Leu Ser Ser Gly Met Val Leu Val Phe Ser Ile	
	305 310 315 320	
60	tgg aac gac aac agc cag tac atg aac tgg ctc gac agc ggc aac gcc	1008
	Trp Asn Asp Asn Ser Gln Tyr Met Asn Trp Leu Asp Ser Gly Asn Ala	
	325 330 335	
65	ggc ccc tgc agc agc acc gag ggc aac cca tcc aac acc ctg gcc aac	1056
	Gly Pro Cys Ser Ser Thr Glu Gly Asn Pro Ser Asn Thr Leu Ala Asn	
	340 345 350	
70	aac ccc aac acg cac gtc gtc ttc tcc aac atc cgc tgg gga gac att	1104

Asn Pro Asn Thr His Val Val Phe Ser Asn Ile Arg Trp Gly Asp Ile  
                   355                                  360                                  365

5    ggg tct act acg aac tcg act gcg ccc ccg ccc ccg cct gcg tcc agc    1152  
      Gly Ser Thr Thr Asn Ser Thr Ala Pro Pro Pro Pro Pro Ala Ser Ser  
                   370                                  375                                  380

10    acg acg ttt tcg act aca cgg agg agc tcg acg act tcg agc agc ccg    1200  
      Thr Thr Phe Ser Thr Thr Arg Arg Ser Ser Thr Thr Ser Ser Ser Pro  
                   385                                  390                                  395                                  400

15    agc tgc acg cag act cac tgg ggg cag tgc ggt ggc att ggg tac agc    1248  
      Ser Cys Thr Gln Thr His Trp Gly Gln Cys Gly Gly Ile Gly Tyr Ser  
                                   405                                  410                                  415

15    ggg tgc aag acg tgc acg tcg ggc act acg tgc cag tat agc aac gac    1296  
      Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys Gln Tyr Ser Asn Asp  
                                   420                                  425                                  430

20    tac tac tcg caa tgc ctt tag    1317  
      Tyr Tyr Ser Gln Cys Leu  
                   435

25    <210> 84  
      <211> 438  
      <212> PRT  
      <213> Trichoderma reesei

30    <400> 84

Met Gln Gln Pro Gly Thr Ser Thr Pro Glu Val His Pro Lys Leu Thr  
   1                                  5                                  10                                  15

35    Thr Tyr Lys Cys Thr Lys Ser Gly Gly Cys Val Ala Gln Asp Thr Ser  
                   20                                  25                                  30

40    Val Val Leu Asp Trp Asn Tyr Arg Trp Met His Asp Ala Asn Tyr Asn  
                   35                                  40                                  45

45    Ser Cys Thr Val Asn Gly Gly Val Asn Thr Thr Leu Cys Pro Asp Glu  
                   50                                  55                                  60

50    Ala Thr Cys Gly Lys Asn Cys Phe Ile Glu Gly Val Asp Tyr Ala Ala  
      65                                  70                                  75                                  80

50    Ser Gly Val Thr Thr Ser Gly Ser Ser Leu Thr Met Asn Gln Tyr Met  
                                   85                                  90                                  95

55    Pro Ser Ser Ser Gly Gly Tyr Ser Ser Val Ser Pro Arg Leu Tyr Leu  
                   100                                  105                                  110



261

Gly Pro Cys Ser Ser Thr Glu Gly Asn Pro Ser Asn Thr Leu Ala Asn  
                     340                    345                    350

5 Asn Pro Asn Thr His Val Val Phe Ser Asn Ile Arg Trp Gly Asp Ile  
                     355                    360                    365

10 Gly Ser Thr Thr Asn Ser Thr Ala Pro Pro Pro Pro Pro Ala Ser Ser  
                     370                    375                    380

15 Thr Thr Phe Ser Thr Thr Arg Arg Ser Ser Thr Thr Ser Ser Ser Pro  
                     385                    390                    395                    400

20 Ser Cys Thr Gln Thr His Trp Gly Gln Cys Gly Gly Ile Gly Tyr Ser  
                     405                    410                    415

Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys Gln Tyr Ser Asn Asp  
                     420                    425                    430

25 Tyr Tyr Ser Gln Cys Leu  
                     435

30 <210> 85  
       <211> 954  
       <212> DNA  
       <213> Artificial Sequence

35 <220>  
       <223> 6GP1

40 <220>  
       <221> CDS  
       <222> (1)..(954)  
       <223> 6GP1

45 <400> 85  
     atg ggc gtg gac ccg ttc gag cgc aac aag atc ctc ggc cgc ggc atc 48  
     Met Gly Val Asp Pro Phe Glu Arg Asn Lys Ile Leu Gly Arg Gly Ile  
     1                    5                    10                    15

50 aac atc ggc aac gcc ctg gag gcc ccg aac gag ggc gac tgg ggc gtg 96  
     Asn Ile Gly Asn Ala Leu Glu Ala Pro Asn Glu Gly Asp Trp Gly Val  
                     20                    25                    30

55 gtg atc aag gac gag ttc ttc gac atc atc aag gag gcc ggc ttc tcc 144  
     Val Ile Lys Asp Glu Phe Phe Asp Ile Ile Lys Glu Ala Gly Phe Ser  
                     35                    40                    45

cac gtg cgc atc ccg atc cgc tgg tcc acc cac gcc tac gcc ttc ccg 192

	His	Val	Arg	Ile	Pro	Ile	Arg	Trp	Ser	Thr	His	Ala	Tyr	Ala	Phe	Pro	
	50					55					60						
5	ccg	tac	aag	atc	atg	gac	cgc	ttc	ttc	aag	cgc	gtg	gac	gag	gtg	atc	240
	Pro	Tyr	Lys	Ile	Met	Asp	Arg	Phe	Phe	Lys	Arg	Val	Asp	Glu	Val	Ile	
	65					70				75					80		
10	aac	ggc	gcc	ctc	aag	cgc	ggc	ctc	gcc	gtg	gcc	atc	aac	atc	cac	cac	288
	Asn	Gly	Ala	Leu	Lys	Arg	Gly	Leu	Ala	Val	Ala	Ile	Asn	Ile	His	His	
					85				90					95			
15	tac	gag	gag	ctc	atg	aac	gac	ccg	gag	gag	cac	aag	gag	cgc	ttc	ctc	336
	Tyr	Glu	Glu	Leu	Met	Asn	Asp	Pro	Glu	Glu	His	Lys	Glu	Arg	Phe	Leu	
				100				105						110			
20	gcc	ctc	tgg	aag	cag	atc	gcc	gac	cgc	tac	aag	gac	tac	ccg	gag	acc	384
	Ala	Leu	Trp	Lys	Gln	Ile	Ala	Asp	Arg	Tyr	Lys	Asp	Tyr	Pro	Glu	Thr	
			115				120						125				
25	ctc	ttc	ttc	gag	atc	ctc	aac	gag	ccg	cac	ggc	aac	ctc	acc	ccg	gag	432
	Leu	Phe	Phe	Glu	Ile	Leu	Asn	Glu	Pro	His	Gly	Asn	Leu	Thr	Pro	Glu	
				130			135					140					
30	aag	tgg	aac	gag	ctg	ctc	gag	gag	gcc	ctc	aag	gtg	atc	cgc	tcc	atc	480
	Lys	Trp	Asn	Glu	Leu	Leu	Glu	Glu	Ala	Leu	Lys	Val	Ile	Arg	Ser	Ile	
	145					150					155					160	
35	gac	aag	aag	cac	acc	atc	atc	att	ggc	acc	gca	gag	tgg	gga	ggc	atc	528
	Asp	Lys	Lys	His	Thr	Ile	Ile	Ile	Gly	Thr	Ala	Glu	Trp	Gly	Gly	Ile	
					165				170						175		
40	tcc	gcc	ctc	gag	aag	ctc	tcc	gtg	ccg	aag	tgg	gag	aag	aat	tcc	atc	576
	Ser	Ala	Leu	Glu	Lys	Leu	Ser	Val	Pro	Lys	Trp	Glu	Lys	Asn	Ser	Ile	
				180					185					190			
45	gtg	acc	atc	cac	tac	tac	aac	ccg	ttc	gag	ttc	acg	cac	cag	ggc	gcc	624
	Val	Thr	Ile	His	Tyr	Tyr	Asn	Pro	Phe	Glu	Phe	Thr	His	Gln	Gly	Ala	
				195			200						205				
50	gag	tgg	gtg	gag	ggc	tcc	gag	aag	tgg	ctt	ggc	cgc	aag	tgg	ggc	tcc	672
	Glu	Trp	Val	Glu	Gly	Ser	Glu	Lys	Trp	Leu	Gly	Arg	Lys	Trp	Gly	Ser	
				210			215					220					
55	ccg	gac	gac	cag	aag	cac	ctc	atc	gag	gag	ttc	aac	ttc	atc	gag	gag	720
	Pro	Asp	Asp	Gln	Lys	His	Leu	Ile	Glu	Glu	Phe	Asn	Phe	Ile	Glu	Glu	
						230					235				240		
60	tgg	tcc	aag	aag	aac	aag	cgc	ccg	atc	tac	atc	ggc	gag	ttt	ggc	gcc	768
	Trp	Ser	Lys	Lys	Asn	Lys	Arg	Pro	Ile	Tyr	Ile	Gly	Glu	Phe	Gly	Ala	
					245				250						255		
65	tac	cgc	aag	gcc	gac	ctc	gag	tcc	cgc	atc	aag	tgg	acc	tcc	ttc	gtg	816
	Tyr	Arg	Lys	Ala	Asp	Leu	Glu	Ser	Arg	Ile	Lys	Trp	Thr	Ser	Phe	Val	
				260				265						270			
70	gtg	cgt	gag	atg	gag	aag	cgc	cgc	tgg	tcc	tgg	gcc	tac	tgg	gag	ttc	864
	Val	Arg	Glu	Met	Glu	Lys	Arg	Arg	Trp	Ser	Trp	Ala	Tyr	Trp	Glu	Phe	

275                      280                      285  
 tgc tcc ggc ttc ggc gtg tac gac acc ctc cgc aag acc tgg aac aag                      912  
 Cys Ser Gly Phe Gly Val Tyr Asp Thr Leu Arg Lys Thr Trp Asn Lys  
 5                      290                      295                      300  
  
 gac ctc ctc gag gcc ctc atc ggc ggc gac tcc atc gag tag                      954  
 Asp Leu Leu Glu Ala Leu Ile Gly Gly Asp Ser Ile Glu  
 10                      305                      310                      315  
  
 <210> 86  
 <211> 317  
 <212> PRT  
 15 <213> Artificial Sequence  
  
 <220>  
 <223> Synthetic Construct  
  
 20 <400> 86  
  
 Met Gly Val Asp Pro Phe Glu Arg Asn Lys Ile Leu Gly Arg Gly Ile  
 1                      5                      10                      15  
  
 25 Asn Ile Gly Asn Ala Leu Glu Ala Pro Asn Glu Gly Asp Trp Gly Val  
                     20                      25                      30  
  
 30 Val Ile Lys Asp Glu Phe Phe Asp Ile Ile Lys Glu Ala Gly Phe Ser  
                     35                      40                      45  
  
 35 His Val Arg Ile Pro Ile Arg Trp Ser Thr His Ala Tyr Ala Phe Pro  
                     50                      55                      60  
  
 Pro Tyr Lys Ile Met Asp Arg Phe Phe Lys Arg Val Asp Glu Val Ile  
 40 65                      70                      75                      80  
  
 Asn Gly Ala Leu Lys Arg Gly Leu Ala Val Ala Ile Asn Ile His His  
                     85                      90                      95  
  
 45 Tyr Glu Glu Leu Met Asn Asp Pro Glu Glu His Lys Glu Arg Phe Leu  
                     100                      105                      110  
  
 50 Ala Leu Trp Lys Gln Ile Ala Asp Arg Tyr Lys Asp Tyr Pro Glu Thr  
                     115                      120                      125  
  
 55 Leu Phe Phe Glu Ile Leu Asn Glu Pro His Gly Asn Leu Thr Pro Glu  
                     130                      135                      140

Lys Trp Asn Glu Leu Leu Glu Glu Ala Leu Lys Val Ile Arg Ser Ile  
 145 150 155 160

5 Asp Lys Lys His Thr Ile Ile Ile Gly Thr Ala Glu Trp Gly Gly Ile  
 165 170 175

10 Ser Ala Leu Glu Lys Leu Ser Val Pro Lys Trp Glu Lys Asn Ser Ile  
 180 185 190

15 Val Thr Ile His Tyr Tyr Asn Pro Phe Glu Phe Thr His Gln Gly Ala  
 195 200 205

Glu Trp Val Glu Gly Ser Glu Lys Trp Leu Gly Arg Lys Trp Gly Ser  
 210 215 220

20 Pro Asp Asp Gln Lys His Leu Ile Glu Glu Phe Asn Phe Ile Glu Glu  
 225 230 235 240

25 Trp Ser Lys Lys Asn Lys Arg Pro Ile Tyr Ile Gly Glu Phe Gly Ala  
 245 250 255

30 Tyr Arg Lys Ala Asp Leu Glu Ser Arg Ile Lys Trp Thr Ser Phe Val  
 260 265 270

35 Val Arg Glu Met Glu Lys Arg Arg Trp Ser Trp Ala Tyr Trp Glu Phe  
 275 280 285

Cys Ser Gly Phe Gly Val Tyr Asp Thr Leu Arg Lys Thr Trp Asn Lys  
 290 295 300

40 Asp Leu Leu Glu Ala Leu Ile Gly Gly Asp Ser Ile Glu  
 305 310 315

45 <210> 87  
 <211> 1248  
 <212> DNA  
 <213> Hordeum vulgare

50 <220>  
 <221> CDS  
 <222> (1)..(1248)  
 <223> Barley AmyI amylase

55 <400> 87  
 atg gca cac caa gtc ctc ttt cag ggg ttc aac tgg gag tcg tgg aag 48

	Met	Ala	His	Gln	Val	Leu	Phe	Gln	Gly	Phe	Asn	Trp	Glu	Ser	Trp	Lys	
	1				5					10					15		
5	cag	agc	ggc	ggg	tgg	tac	aac	atg	atg	ggc	aag	gtc	gac	gac	atc		96
	Gln	Ser	Gly	Gly	Trp	Tyr	Asn	Met	Met	Gly	Lys	Val	Asp	Asp	Ile		
				20					25					30			
10	gcc	gct	gcc	gga	gtc	acc	cac	gtc	tgg	ctg	cca	ccg	ccg	tcg	cac	tcc	144
	Ala	Ala	Ala	Gly	Val	Thr	His	Val	Trp	Leu	Pro	Pro	Pro	Ser	His	Ser	
				35					40					45			
15	gtc	tcc	aac	gaa	ggc	tac	atg	cct	ggc	cgg	ctg	tac	gac	atc	gac	gcg	192
	Val	Ser	Asn	Glu	Gly	Tyr	Met	Pro	Gly	Arg	Leu	Tyr	Asp	Ile	Asp	Ala	
			50				55						60				
20	tcc	aag	tac	ggc	aac	gcg	gcg	gag	ctc	aag	tcg	ctc	atc	ggc	gcg	ctc	240
	Ser	Lys	Tyr	Gly	Asn	Ala	Glu	Leu	Lys	Ser	Leu	Ile	Gly	Ala	Leu		
	65					70				75					80		
25	cac	ggc	aag	ggc	gtg	cag	gcc	atc	gcc	gac	atc	gtc	atc	aac	cac	cgc	288
	His	Gly	Lys	Gly	Val	Gln	Ala	Ile	Ala	Asp	Ile	Val	Ile	Asn	His	Arg	
					85					90					95		
30	tgc	gcc	gac	tac	aag	gat	agc	cgc	ggc	atc	tac	tgc	atc	ttc	gag	ggc	336
	Cys	Ala	Asp	Tyr	Lys	Asp	Ser	Arg	Gly	Ile	Tyr	Cys	Ile	Phe	Glu	Gly	
				100					105						110		
35	ggc	acc	tcc	gac	ggc	cgc	ctc	gac	tgg	ggc	ccc	cac	atg	atc	tgt	cgc	384
	Gly	Thr	Ser	Asp	Gly	Arg	Leu	Asp	Trp	Gly	Pro	His	Met	Ile	Cys	Arg	
				115					120					125			
40	gac	gac	acc	aaa	tac	tcc	gat	ggc	acc	gca	aac	ctc	gac	acc	gga	gcc	432
	Asp	Asp	Thr	Lys	Tyr	Ser	Asp	Gly	Thr	Ala	Asn	Leu	Asp	Thr	Gly	Ala	
				130					135					140			
45	gac	ttc	gcc	gcc	gcg	ccc	gac	atc	gac	cac	ctc	aac	gac	cgg	gtc	cag	480
	Asp	Phe	Ala	Ala	Ala	Pro	Asp	Ile	Asp	His	Leu	Asn	Asp	Arg	Val	Gln	
						150					155				160		
50	cgc	gag	ctc	aag	gag	tgg	ctc	ctc	tgg	ctc	aag	agc	gac	ctc	ggc	ttc	528
	Arg	Glu	Leu	Lys	Glu	Trp	Leu	Leu	Trp	Leu	Lys	Ser	Asp	Leu	Gly	Phe	
					165					170					175		
55	gac	gcg	tgg	cgc	ctt	gac	ttc	gcc	agg	ggc	tac	tcg	ccg	gag	atg	gcc	576
	Asp	Ala	Trp	Arg	Leu	Asp	Phe	Ala	Arg	Gly	Tyr	Ser	Pro	Glu	Met	Ala	
				180					185						190		
60	aag	gtg	tac	atc	gac	ggc	aca	tcc	ccg	agc	ctc	gcc	gtg	gcc	gag	gtg	624
	Lys	Val	Tyr	Ile	Asp	Gly	Thr	Ser	Pro	Ser	Leu	Ala	Val	Ala	Glu	Val	
				195					200					205			
65	tgg	gac	aat	atg	gcc	acc	ggc	ggc	gac	ggc	aag	ccc	aac	tac	gac	cag	672
	Trp	Asp	Asn	Met	Ala	Thr	Gly	Gly	Asp	Gly	Lys	Pro	Asn	Tyr	Asp	Gln	
				210				215					220				
70	gac	gcg	cac	cgg	cag	aat	ctg	gtg	aac	tgg	gtg	gac	aag	gtg	ggc	ggc	720
	Asp	Ala	His	Arg	Gln	Asn	Leu	Val	Asn	Trp	Val	Asp	Lys	Val	Gly	Gly	

	225		230		235		240	
	gcg gcc tcg gca ggc atg gtg ttc gac ttc acg acc aaa ggg ata ctg							768
5	Ala Ala Ser Ala Gly Met Val Phe Asp Phe Thr Thr Lys Gly Ile Leu	245		250		255		
	aac gct gcc gtg gag ggc gag ctg tgg agg ctg atc gac ccg cag ggg							816
10	Asn Ala Ala Val Glu Gly Glu Leu Trp Arg Leu Ile Asp Pro Gln Gly	260		265		270		
	aag gcc ccc ggc gtg atg gga tgg tgg ccg gcc aag gcc gtc acc ttc							864
15	Lys Ala Pro Gly Val Met Gly Trp Trp Pro Ala Lys Ala Val Thr Phe	275		280		285		
	gtc gac aac cac gat aca ggc tcc acg cag gcc atg tgg cca ttc ccc							912
20	Val Asp Asn His Asp Thr Gly Ser Thr Gln Ala Met Trp Pro Phe Pro	290		295		300		
	tcc gac aag gtc atg cag ggc tac gcg tac atc ctc acc cac ccc ggc							960
25	Ser Asp Lys Val Met Gln Gly Tyr Ala Tyr Ile Leu Thr His Pro Gly	305		310		315		320
	atc cca tgc atc ttc tac gac cat ttc ttc aac tgg ggg ttt aag gac							1008
30	Ile Pro Cys Ile Phe Tyr Asp His Phe Phe Asn Trp Gly Phe Lys Asp	325		330		335		
	cag atc gcg gcg ctg gtg gcg atc agg aag cgc aac ggc atc acg gcg							1056
35	Gln Ile Ala Ala Leu Val Ala Ile Arg Lys Arg Asn Gly Ile Thr Ala	340		345		350		
	acg agc gct ctg aag atc ctc atg cac gaa gga gat gcc tac gtc gcc							1104
40	Thr Ser Ala Leu Lys Ile Leu Met His Glu Gly Asp Ala Tyr Val Ala	355		360		365		
	gag ata gac ggc aag gtg gtg gtg aag atc ggg tcc agg tac gac gtc							1152
45	Glu Ile Asp Gly Lys Val Val Val Lys Ile Gly Ser Arg Tyr Asp Val	370		375		380		
	ggg gcg gtg atc ccg gcc ggg ttc gtg acc tcg gca cac ggc aac gac							1200
50	Gly Ala Val Ile Pro Ala Gly Phe Val Thr Ser Ala His Gly Asn Asp	385		390		395		400
	tac gcc gtc tgg gag aag aac ggt gcc gcg gca aca cta caa cgg agc							1248
55	Tyr Ala Val Trp Glu Lys Asn Gly Ala Ala Ala Thr Leu Gln Arg Ser	405		410		415		
	<210> 88							
	<211> 416							
	<212> PRT							
	<213> Hordeum vulgare							
	<400> 88							
	Met Ala His Gln Val Leu Phe Gln Gly Phe Asn Trp Glu Ser Trp Lys							
	1 5 10 15							

Gln Ser Gly Gly Trp Tyr Asn Met Met Met Gly Lys Val Asp Asp Ile  
                   20                  25                  30  
 5  
 Ala Ala Ala Gly Val Thr His Val Trp Leu Pro Pro Pro Ser His Ser  
           35                  40                  45  
 10 Val Ser Asn Glu Gly Tyr Met Pro Gly Arg Leu Tyr Asp Ile Asp Ala  
       50                  55                  60  
 15 Ser Lys Tyr Gly Asn Ala Ala Glu Leu Lys Ser Leu Ile Gly Ala Leu  
       65                  70                  75                  80  
 20 His Gly Lys Gly Val Gln Ala Ile Ala Asp Ile Val Ile Asn His Arg  
                   85                  90                  95  
 Cys Ala Asp Tyr Lys Asp Ser Arg Gly Ile Tyr Cys Ile Phe Glu Gly  
           100                  105                  110  
 25 Gly Thr Ser Asp Gly Arg Leu Asp Trp Gly Pro His Met Ile Cys Arg  
           115                  120                  125  
 30 Asp Asp Thr Lys Tyr Ser Asp Gly Thr Ala Asn Leu Asp Thr Gly Ala  
       130                  135                  140  
 35 Asp Phe Ala Ala Ala Pro Asp Ile Asp His Leu Asn Asp Arg Val Gln  
       145                  150                  155                  160  
 40 Arg Glu Leu Lys Glu Trp Leu Leu Trp Leu Lys Ser Asp Leu Gly Phe  
                   165                  170                  175  
 Asp Ala Trp Arg Leu Asp Phe Ala Arg Gly Tyr Ser Pro Glu Met Ala  
           180                  185                  190  
 45 Lys Val Tyr Ile Asp Gly Thr Ser Pro Ser Leu Ala Val Ala Glu Val  
           195                  200                  205  
 50 Trp Asp Asn Met Ala Thr Gly Gly Asp Gly Lys Pro Asn Tyr Asp Gln  
       210                  215                  220  
 55 Asp Ala His Arg Gln Asn Leu Val Asn Trp Val Asp Lys Val Gly Gly  
       225                  230                  235                  240



Ala Ala Ser Ala Gly Met Val Phe Asp Phe Thr Thr Lys Gly Ile Leu  
245 250 255

5 Asn Ala Ala Val Glu Gly Glu Leu Trp Arg Leu Ile Asp Pro Gln Gly  
260 265 270

10 Lys Ala Pro Gly Val Met Gly Trp Trp Pro Ala Lys Ala Val Thr Phe  
275 280 285

15 Val Asp Asn His Asp Thr Gly Ser Thr Gln Ala Met Trp Pro Phe Pro  
290 295 300

Ser Asp Lys Val Met Gln Gly Tyr Ala Tyr Ile Leu Thr His Pro Gly  
305 310 315 320

20 Ile Pro Cys Ile Phe Tyr Asp His Phe Phe Asn Trp Gly Phe Lys Asp  
325 330 335

25 Gln Ile Ala Ala Leu Val Ala Ile Arg Lys Arg Asn Gly Ile Thr Ala  
340 345 350

30 Thr Ser Ala Leu Lys Ile Leu Met His Glu Gly Asp Ala Tyr Val Ala  
355 360 365

35 Glu Ile Asp Gly Lys Val Val Val Lys Ile Gly Ser Arg Tyr Asp Val  
370 375 380

Gly Ala Val Ile Pro Ala Gly Phe Val Thr Ser Ala His Gly Asn Asp  
385 390 395 400

40 Tyr Ala Val Trp Glu Lys Asn Gly Ala Ala Ala Thr Leu Gln Arg Ser  
405 410 415

45 <210> 89  
<211> 1401  
<212> DNA  
<213> Artificial Sequence

50 <220>  
<223> Trichoderma reesei  $\beta$ -Glucosidase 2

55 <220>  
<221> CDS  
<222> (1)..(1401)  
<223> Trichoderma reesei  $\beta$ -Glucosidase 2

<400> 89		
	atg ttg ccc aag gac ttt cag tgg ggg ttc gcc acg gct gcc tac cag	48
5	Met Leu Pro Lys Asp Phe Gln Trp Gly Phe Ala Thr Ala Ala Tyr Gln	
	1 5 10 15	
	atc gag ggc gcc gtc gac cag gac ggc cgc ggc ccc agc atc tgg gac	96
	Ile Glu Gly Ala Val Asp Gln Asp Gly Arg Gly Pro Ser Ile Trp Asp	
	20 25 30	
10	acg ttc tgc gcg cag ccc ggc aag atc gcc gac ggc tcg tcg ggc gtg	144
	Thr Phe Cys Ala Gln Pro Gly Lys Ile Ala Asp Gly Ser Ser Gly Val	
	35 40 45	
15	acg gcg tgc gac tcg tac aac cgc acg gcc gag gac att gcg ctg ctg	192
	Thr Ala Cys Asp Ser Tyr Asn Arg Thr Ala Glu Asp Ile Ala Leu Leu	
	50 55 60	
20	aag tcg ctc ggg gcc aag agc tac cgc ttc tcc atc tcg tgg tcg cgc	240
	Lys Ser Leu Gly Ala Lys Ser Tyr Arg Phe Ser Ile Ser Trp Ser Arg	
	65 70 75 80	
	atc atc ccc gag ggc ggc cgc ggc gat gcc gtc aac cag gcg ggc atc	288
	Ile Ile Pro Glu Gly Gly Arg Gly Asp Ala Val Asn Gln Ala Gly Ile	
25	85 90 95	
	gac cac tac gtc aag ttc gtc gac gac ctg ctc gac gcc ggc atc acg	336
	Asp His Tyr Val Lys Phe Val Asp Asp Leu Leu Asp Ala Gly Ile Thr	
	100 105 110	
30	ccc ttc atc acc ctc ttc cac tgg gac ctg ccc gag ggc ctg cat cag	384
	Pro Phe Ile Thr Leu Phe His Trp Asp Leu Pro Glu Gly Leu His Gln	
	115 120 125	
35	cgg tac ggg ggg ctg ctg aac cgc acc gag ttc ccg ctc gac ttt gaa	432
	Arg Tyr Gly Gly Leu Leu Asn Arg Thr Glu Phe Pro Leu Asp Phe Glu	
	130 135 140	
40	aac tac gcc cgc gtc atg ttc agg gcg ctg ccc aag gtg cgc aac tgg	480
	Asn Tyr Ala Arg Val Met Phe Arg Ala Leu Pro Lys Val Arg Asn Trp	
	145 150 155 160	
	atc acc ttc aac gag ccg ctg tgc tcg gcc atc ccg ggc tac ggc tcc	528
	Ile Thr Phe Asn Glu Pro Leu Cys Ser Ala Ile Pro Gly Tyr Gly Ser	
45	165 170 175	
	ggc acc ttc gcc ccc ggc cgg cag agc acc tcg gag ccg tgg acc gtc	576
	Gly Thr Phe Ala Pro Gly Arg Gln Ser Thr Ser Glu Pro Trp Thr Val	
	180 185 190	
50	ggc cac aac atc ctc gtc gcc cac ggc cgc gcc gtc aag gcg tac cgc	624
	Gly His Asn Ile Leu Val Ala His Gly Arg Ala Val Lys Ala Tyr Arg	
	195 200 205	
55	gac gac ttc aag ccc gcc agc ggc gac ggc cag atc ggc atc gtc ctc	672
	Asp Asp Phe Lys Pro Ala Ser Gly Asp Gly Gln Ile Gly Ile Val Leu	
	210 215 220	

	aac ggc gac ttc acc tac ccc tgg gac gcc gcc gac ccg gcc gac aag	720
	Asn Gly Asp Phe Thr Tyr Pro Trp Asp Ala Ala Asp Pro Ala Asp Lys	
	225 230 235 240	
5		
	gag gcg gcc gag cgg cgc ctc gag ttc ttc acg gcc tgg ttc gcg gac	768
	Glu Ala Ala Glu Arg Arg Leu Glu Phe Phe Thr Ala Trp Phe Ala Asp	
	245 250 255	
10		
	ccc atc tac ttg ggc gac tac ccg gcg tcg atg cgc aag cag ctg ggc	816
	Pro Ile Tyr Leu Gly Asp Tyr Pro Ala Ser Met Arg Lys Gln Leu Gly	
	260 265 270	
15		
	gac cgg ctg ccg acc ttt acg ccc gag gag cgc gcc ctc gtc cac ggc	864
	Asp Arg Leu Pro Thr Phe Thr Pro Glu Glu Arg Ala Leu Val His Gly	
	275 280 285	
20		
	tcc aac gac ttt tac ggc atg aac cac tac acg tcc aac tac atc cgc	912
	Ser Asn Asp Phe Tyr Gly Met Asn His Tyr Thr Ser Asn Tyr Ile Arg	
	290 295 300	
25		
	cac cgc agc tcg ccc gcc tcc gcc gac gac acc gtc ggc aac gtc gac	960
	His Arg Ser Ser Pro Ala Ser Ala Asp Asp Thr Val Gly Asn Val Asp	
	305 310 315 320	
30		
	gtg ctc ttc acc aac aag cag ggc aac tgc atc ggc ccc gag acg cag	1008
	Val Leu Phe Thr Asn Lys Gln Gly Asn Cys Ile Gly Pro Glu Thr Gln	
	325 330 335	
35		
	tcc ccc tgg ctg cgc ccc tgt gcc gcc ggc ttc cgc gac ttc ctg gtg	1056
	Ser Pro Trp Leu Arg Pro Cys Ala Ala Gly Phe Arg Asp Phe Leu Val	
	340 345 350	
40		
	tgg atc agc aag agg tac ggc tac ccg ccc atc tac gtg acg gag aac	1104
	Trp Ile Ser Lys Arg Tyr Gly Tyr Pro Pro Ile Tyr Val Thr Glu Asn	
	355 360 365	
45		
	ggc acg agc atc aag ggc gag agc gac ttg ccc aag gag aag att ctc	1152
	Gly Thr Ser Ile Lys Gly Glu Ser Asp Leu Pro Lys Glu Lys Ile Leu	
	370 375 380	
50		
	gaa gat gac ttc agg gtc aag tac tat aac gag tac atc cgt gcc atg	1200
	Glu Asp Asp Phe Arg Val Lys Tyr Tyr Asn Glu Tyr Ile Arg Ala Met	
	385 390 395 400	
55		
	gtt acc gcc gtg gag ctg gac ggg gtc aac gtc aag ggg tac ttt gcc	1248
	Val Thr Ala Val Glu Leu Asp Gly Val Asn Val Lys Gly Tyr Phe Ala	
	405 410 415	
60		
	tgg tcg ctc atg gac aac ttt gag tgg gcg gac ggc tac gtg acg agg	1296
	Trp Ser Leu Met Asp Asn Phe Glu Trp Ala Asp Gly Tyr Val Thr Arg	
	420 425 430	
65		
	ttt ggg gtt acg tat gtg gat tat gag aat ggg cag aag cgg ttc ccc	1344
	Phe Gly Val Thr Tyr Val Asp Tyr Glu Asn Gly Gln Lys Arg Phe Pro	
	435 440 445	

```

aag aag agc gca aag agc ttg aag ccg ctg ttt gac gag ctg att gcg      1392
Lys Lys Ser Ala Lys Ser Leu Lys Pro Leu Phe Asp Glu Leu Ile Ala
    450                455                460

5   gcg gcg tga                      1401
    Ala Ala
    465

10  <210>  90
    <211>  466
    <212>  PRT
    <213>  Artificial Sequence

15  <220>
    <223>  Synthetic Construct

    <400>  90

20  Met Leu Pro Lys Asp Phe Gln Trp Gly Phe Ala Thr Ala Ala Tyr Gln
    1              5              10              15

    Ile Glu Gly Ala Val Asp Gln Asp Gly Arg Gly Pro Ser Ile Trp Asp
25              20              25              30

    Thr Phe Cys Ala Gln Pro Gly Lys Ile Ala Asp Gly Ser Ser Gly Val
    35              40              45

30  Thr Ala Cys Asp Ser Tyr Asn Arg Thr Ala Glu Asp Ile Ala Leu Leu
    50              55              60

35  Lys Ser Leu Gly Ala Lys Ser Tyr Arg Phe Ser Ile Ser Trp Ser Arg
    65              70              75              80

40  Ile Ile Pro Glu Gly Gly Arg Gly Asp Ala Val Asn Gln Ala Gly Ile
    85              90              95

    Asp His Tyr Val Lys Phe Val Asp Asp Leu Leu Asp Ala Gly Ile Thr
45              100              105              110

    Pro Phe Ile Thr Leu Phe His Trp Asp Leu Pro Glu Gly Leu His Gln
    115              120              125

50  Arg Tyr Gly Gly Leu Leu Asn Arg Thr Glu Phe Pro Leu Asp Phe Glu
    130              135              140

55  Asn Tyr Ala Arg Val Met Phe Arg Ala Leu Pro Lys Val Arg Asn Trp
    145              150              155              160

```

	Ile	Thr	Phe	Asn	Glu	Pro	Leu	Cys	Ser	Ala	Ile	Pro	Gly	Tyr	Gly	Ser	
					165					170						175	
5																	
	Gly	Thr	Phe	Ala	Pro	Gly	Arg	Gln	Ser	Thr	Ser	Glu	Pro	Trp	Thr	Val	
				180					185					190			
10																	
	Gly	His	Asn	Ile	Leu	Val	Ala	His	Gly	Arg	Ala	Val	Lys	Ala	Tyr	Arg	
			195					200					205				
15																	
	Asp	Asp	Phe	Lys	Pro	Ala	Ser	Gly	Asp	Gly	Gln	Ile	Gly	Ile	Val	Leu	
		210					215					220					
20																	
	Asn	Gly	Asp	Phe	Thr	Tyr	Pro	Trp	Asp	Ala	Ala	Asp	Pro	Ala	Asp	Lys	
	225					230					235					240	
25																	
	Glu	Ala	Ala	Glu	Arg	Arg	Leu	Glu	Phe	Phe	Thr	Ala	Trp	Phe	Ala	Asp	
					245					250					255		
30																	
	Pro	Ile	Tyr	Leu	Gly	Asp	Tyr	Pro	Ala	Ser	Met	Arg	Lys	Gln	Leu	Gly	
				260					265					270			
35																	
	Asp	Arg	Leu	Pro	Thr	Phe	Thr	Pro	Glu	Glu	Arg	Ala	Leu	Val	His	Gly	
			275					280					285				
40																	
	Ser	Asn	Asp	Phe	Tyr	Gly	Met	Asn	His	Tyr	Thr	Ser	Asn	Tyr	Ile	Arg	
		290					295					300					
45																	
	His	Arg	Ser	Ser	Pro	Ala	Ser	Ala	Asp	Asp	Thr	Val	Gly	Asn	Val	Asp	
	305					310					315					320	
50																	
	Val	Leu	Phe	Thr	Asn	Lys	Gln	Gly	Asn	Cys	Ile	Gly	Pro	Glu	Thr	Gln	
					325					330					335		
55																	
	Ser	Pro	Trp	Leu	Arg	Pro	Cys	Ala	Ala	Gly	Phe	Arg	Asp	Phe	Leu	Val	
				340					345					350			
60																	
	Trp	Ile	Ser	Lys	Arg	Tyr	Gly	Tyr	Pro	Pro	Ile	Tyr	Val	Thr	Glu	Asn	
			355					360					365				
65																	
	Gly	Thr	Ser	Ile	Lys	Gly	Glu	Ser	Asp	Leu	Pro	Lys	Glu	Lys	Ile	Leu	
		370					375					380					

Glu Asp Asp Phe Arg Val Lys Tyr Tyr Asn Glu Tyr Ile Arg Ala Met  
 385 390 395 400  
 5 Val Thr Ala Val Glu Leu Asp Gly Val Asn Val Lys Gly Tyr Phe Ala  
 405 410 415  
 10 Trp Ser Leu Met Asp Asn Phe Glu Trp Ala Asp Gly Tyr Val Thr Arg  
 420 425 430  
 15 Phe Gly Val Thr Tyr Val Asp Tyr Glu Asn Gly Gln Lys Arg Phe Pro  
 435 440 445  
 20 Lys Lys Ser Ala Lys Ser Leu Lys Pro Leu Phe Asp Glu Leu Ile Ala  
 450 455 460  
 Ala Ala  
 465  
 25 <210> 91  
 <211> 2103  
 <212> DNA  
 <213> Artificial Sequence  
 30 <220>  
 <223> Trichoderma reesei  $\beta$ -Glucosidase D  
 35 <220>  
 <221> CDS  
 <222> (1)..(2103)  
 <223> Trichoderma reesei  $\beta$ -Glucosidase D  
 40 <400> 91  
 atg att ctc ggc tgt gaa agc aca ggt gtc atc tct gcc gtc aaa cac 48  
 Met Ile Leu Gly Cys Glu Ser Thr Gly Val Ile Ser Ala Val Lys His  
 1 5 10 15  
 45 ttt gtc gcc aac gac cag gag cac gag cgg cga gcg gtc gac tgt ctc 96  
 Phe Val Ala Asn Asp Gln Glu His Glu Arg Arg Ala Val Asp Cys Leu  
 20 25 30  
 50 atc acc cag cgg gct ctc cgg gag gtc tat ctg cga ccc ttc cag atc 144  
 Ile Thr Gln Arg Ala Leu Arg Glu Val Tyr Leu Arg Pro Phe Gln Ile  
 35 40 45  
 gta gcc cga gat gca agg ccc ggc gca ttg atg aca tcc tac aac aag 192  
 Val Ala Arg Asp Ala Arg Pro Gly Ala Leu Met Thr Ser Tyr Asn Lys  
 55 50 55 60  
 gtc aat ggc aag cac gtc gct gac agc gcc gag ttc ctt cag ggc att 240

	Val	Asn	Gly	Lys	His	Val	Ala	Asp	Ser	Ala	Glu	Phe	Leu	Gln	Gly	Ile	
	65					70					75					80	
5	ctc	cgg	act	gag	tgg	aat	tgg	gac	cct	ctc	att	gtc	agc	gac	tgg	tac	288
	Leu	Arg	Thr	Glu	Trp	Asn	Trp	Asp	Pro	Leu	Ile	Val	Ser	Asp	Trp	Tyr	
					85					90					95		
10	ggc	acc	tac	acc	act	att	gat	gcc	atc	aaa	gcc	ggc	ctt	gat	ctc	gag	336
	Gly	Thr	Tyr	Thr	Thr	Ile	Asp	Ala	Ile	Lys	Ala	Gly	Leu	Asp	Leu	Glu	
					100				105					110			
15	atg	ccg	ggc	gtt	tca	cga	tat	cgc	ggc	aaa	tac	atc	gag	tct	gct	ctg	384
	Met	Pro	Gly	Val	Ser	Arg	Tyr	Arg	Gly	Lys	Tyr	Ile	Glu	Ser	Ala	Leu	
			115					120					125				
20	cag	gcc	cgt	ttg	ctg	aag	cag	tcc	act	atc	gat	gag	cgc	gct	cgc	cgc	432
	Gln	Ala	Arg	Leu	Leu	Lys	Gln	Ser	Thr	Ile	Asp	Glu	Arg	Ala	Arg	Arg	
			130				135					140					
25	gtg	ctc	agg	ttc	gcc	cag	aag	gcc	agc	cat	ctc	aag	gtc	tcc	gag	gta	480
	Val	Leu	Arg	Phe	Ala	Gln	Lys	Ala	Ser	His	Leu	Lys	Val	Ser	Glu	Val	
						150					155				160		
30	gag	caa	ggc	cgt	gac	ttc	cca	gag	gat	cgc	gtc	ctc	aac	cgt	cag	atc	528
	Glu	Gln	Gly	Arg	Asp	Phe	Pro	Glu	Asp	Arg	Val	Leu	Asn	Arg	Gln	Ile	
					165					170					175		
35	tgc	ggc	agc	agc	att	gtc	cta	ctg	aag	aat	gag	aac	tcc	atc	tta	cct	576
	Cys	Gly	Ser	Ser	Ile	Val	Leu	Leu	Lys	Asn	Glu	Asn	Ser	Ile	Leu	Pro	
					180				185					190			
40	ctc	ccc	aag	tcc	gtc	aag	aag	gtc	gcc	ctt	gtt	ggc	tcc	cac	gtg	cgt	624
	Leu	Pro	Lys	Ser	Val	Lys	Lys	Val	Ala	Leu	Val	Gly	Ser	His	Val	Arg	
			195					200					205				
45	cta	ccg	gct	atc	tcg	gga	gga	ggc	agc	gcc	tct	ctt	gtc	cct	tac	tat	672
	Leu	Pro	Ala	Ile	Ser	Gly	Gly	Gly	Ser	Ala	Ser	Leu	Val	Pro	Tyr	Tyr	
			210				215					220					
50	gcc	ata	tct	cta	tac	gat	gcc	gtc	tct	gag	gta	cta	gcc	ggc	gcc	acg	720
	Ala	Ile	Ser	Leu	Tyr	Asp	Ala	Val	Ser	Glu	Val	Leu	Ala	Gly	Ala	Thr	
						230				235					240		
55	atc	acg	cac	gag	gtc	ggc	tat	gcc	cac	caa	atg	ctg	ccc	gtc	atc		768
	Ile	Thr	His	Glu	Val	Gly	Ala	Tyr	Ala	His	Gln	Met	Leu	Pro	Val	Ile	
					245				250					255			
60	gac	gca	atg	atc	agc	aac	gcc	gta	atc	cac	ttc	tac	aac	gac	ccc	atc	816
	Asp	Ala	Met	Ile	Ser	Asn	Ala	Val	Ile	His	Phe	Tyr	Asn	Asp	Pro	Ile	
					260				265					270			
65	gat	gtc	aaa	gac	aga	aag	ctc	ctt	ggc	agt	gag	aac	gta	tcg	tcg	aca	864
	Asp	Val	Lys	Asp	Arg	Lys	Leu	Leu	Gly	Ser	Glu	Asn	Val	Ser	Ser	Thr	
			275				280						285				
70	tcg	ttc	cag	ctc	atg	gat	tac	aac	aac	atc	cca	acg	ctc	aac	aag	gcc	912
	Ser	Phe	Gln	Leu	Met	Asp	Tyr	Asn	Asn	Ile	Pro	Thr	Leu	Asn	Lys	Ala	

	290	295	300	
	atg ttc tgg ggt act ctc gtg ggc gag ttt atc cct acc gcc acg gga			960
	Met Phe Trp Gly Thr Leu Val Gly Glu Phe Ile Pro Thr Ala Thr Gly			
5	305	310	315	320
	att tgg gaa ttt ggc ctc agt gtc ttt ggc act gcc gac ctt tat att			1008
	Ile Trp Glu Phe Gly Leu Ser Val Phe Gly Thr Ala Asp Leu Tyr Ile			
		325	330	335
10	gat aat gag ctc gtg att gaa aat aca aca cat cag acg cgt gga acc			1056
	Asp Asn Glu Leu Val Ile Glu Asn Thr Thr His Gln Thr Arg Gly Thr			
		340	345	350
15	gcc ttt ttc gga aag gga acg acg gaa aaa gtc gct acc agg agg atg			1104
	Ala Phe Phe Gly Lys Gly Thr Thr Glu Lys Val Ala Thr Arg Arg Met			
		355	360	365
	gtg gcc ggc agc acc tac aag ctg cgt ctc gag ttt ggg tct gcc aac			1152
20	Val Ala Gly Ser Thr Tyr Lys Leu Arg Leu Glu Phe Gly Ser Ala Asn			
		370	375	380
	acg acc aag atg gag acg acc ggt gtt gtc aac ttt ggc ggc ggt gcc			1200
	Thr Thr Lys Met Glu Thr Thr Gly Val Val Asn Phe Gly Gly Gly Ala			
25	385	390	395	400
	gta cac ctg ggt gcc tgt ctc aag gtc gac cca cag gag atg att gcg			1248
	Val His Leu Gly Ala Cys Leu Lys Val Asp Pro Gln Glu Met Ile Ala			
		405	410	415
30	cgg gcc gtc aag gcc gca gcc gat gcc gac tac acc atc atc tgc acg			1296
	Arg Ala Val Lys Ala Ala Ala Asp Ala Asp Tyr Thr Ile Ile Cys Thr			
		420	425	430
35	gga ctc agc ggc gag tgg gag tct gag ggt ttt gac cgg cct cac atg			1344
	Gly Leu Ser Gly Glu Trp Glu Ser Glu Gly Phe Asp Arg Pro His Met			
		435	440	445
	gac ctg ccc cct ggt gtg gac acc atg atc tcg caa gtt ctt gac gcc			1392
40	Asp Leu Pro Pro Gly Val Asp Thr Met Ile Ser Gln Val Leu Asp Ala			
		450	455	460
	gct ccc aat gct gta gtc gtc aac cag tca ggc acc cca gtg aca atg			1440
	Ala Pro Asn Ala Val Val Val Asn Gln Ser Gly Thr Pro Val Thr Met			
45	465	470	475	480
	agc tgg gct cat aaa gca aag gcc att gtg cag gct tgg tat ggt ggt			1488
	Ser Trp Ala His Lys Ala Lys Ala Ile Val Gln Ala Trp Tyr Gly Gly			
		485	490	495
50	aac gag aca ggc cac gga atc tcc gat gtg ctc ttt ggc aac gtc aac			1536
	Asn Glu Thr Gly His Gly Ile Ser Asp Val Leu Phe Gly Asn Val Asn			
		500	505	510
55	ccg tcg ggg aaa ctc tcc cta tcg tgg cca gtc gat gtg aag cac aac			1584
	Pro Ser Gly Lys Leu Ser Leu Ser Trp Pro Val Asp Val Lys His Asn			
		515	520	525



cca gca tat ctc aac tac gcc agc gtt ggt gga cgg gtc ttg tat ggc 1632  
 Pro Ala Tyr Leu Asn Tyr Ala Ser Val Gly Gly Arg Val Leu Tyr Gly  
 530 535 540  
 5 gag gat gtt tac gtt ggc tac aag ttc tac gac aaa acg gag agg gag 1680  
 Glu Asp Val Tyr Val Gly Tyr Lys Phe Tyr Asp Lys Thr Glu Arg Glu  
 545 550 555 560  
 10 gtt ctg ttt cct ttt ggg cat ggc ctg tct tac gct acc ttc aag ctc 1728  
 Val Leu Phe Pro Phe Gly His Gly Leu Ser Tyr Ala Thr Phe Lys Leu  
 565 570 575  
 15 cca gat tct acc gtg agg acg gtc ccc gaa acc ttc cac ccg gac cag 1776  
 Pro Asp Ser Thr Val Arg Thr Val Pro Glu Thr Phe His Pro Asp Gln  
 580 585 590  
 20 ccc aca gta gcc att gtc aag atc aag aac acg agc agt gtc ccg ggc 1824  
 Pro Thr Val Ala Ile Val Lys Ile Lys Asn Thr Ser Ser Val Pro Gly  
 595 600 605  
 25 gcc cag gtc ctg cag tta tac att tcg gcc cca aac tcg cct aca cat 1872  
 Ala Gln Val Leu Gln Leu Ile Ser Ala Pro Asn Ser Pro Thr His  
 610 615 620  
 cgc ccg gtc aag gag ctg cac gga ttc gaa aag gtg tat ctt gaa gct 1920  
 Arg Pro Val Lys Glu Leu His Gly Phe Glu Lys Val Tyr Leu Glu Ala  
 625 630 635 640  
 30 ggc gag gag aag gag gta caa ata ccc att gac cag tac gct act agc 1968  
 Gly Glu Glu Lys Glu Val Gln Ile Pro Ile Asp Gln Tyr Ala Thr Ser  
 645 650 655  
 35 ttc tgg gac gag att gag agc atg tgg aag agc gag agg ggc att tat 2016  
 Phe Trp Asp Glu Ile Glu Ser Met Trp Lys Ser Glu Arg Gly Ile Tyr  
 660 665 670  
 40 gat gtg ctt gta gga ttc tcg agt cag gaa atc tcg ggc aag ggg aag 2064  
 Asp Val Leu Val Gly Phe Ser Ser Gln Glu Ile Ser Gly Lys Gly Lys  
 675 680 685  
 45 ctg att gtg cct gaa acg cga ttc tgg atg ggg ctg tag 2103  
 Leu Ile Val Pro Glu Thr Arg Phe Trp Met Gly Leu  
 690 695 700  
 <210> 92  
 <211> 700  
 <212> PRT  
 50 <213> Artificial Sequence  
 <220>  
 <223> Synthetic Construct  
 55 <400> 92  
 Met Ile Leu Gly Cys Glu Ser Thr Gly Val Ile Ser Ala Val Lys His

1	5	10	15
5	Phe Val Ala Asn Asp Gln Glu His Glu Arg Arg Ala Val Asp Cys Leu	20	25
10	Ile Thr Gln Arg Ala Leu Arg Glu Val Tyr Leu Arg Pro Phe Gln Ile	35	40
15	Val Ala Arg Asp Ala Arg Pro Gly Ala Leu Met Thr Ser Tyr Asn Lys	50	55
20	Val Asn Gly Lys His Val Ala Asp Ser Ala Glu Phe Leu Gln Gly Ile	65	70
25	Leu Arg Thr Glu Trp Asn Trp Asp Pro Leu Ile Val Ser Asp Trp Tyr	85	90
30	Gly Thr Tyr Thr Thr Ile Asp Ala Ile Lys Ala Gly Leu Asp Leu Glu	100	105
35	Met Pro Gly Val Ser Arg Tyr Arg Gly Lys Tyr Ile Glu Ser Ala Leu	115	120
40	Gln Ala Arg Leu Leu Lys Gln Ser Thr Ile Asp Glu Arg Ala Arg Arg	130	135
45	Val Leu Arg Phe Ala Gln Lys Ala Ser His Leu Lys Val Ser Glu Val	145	150
50	Glu Gln Gly Arg Asp Phe Pro Glu Asp Arg Val Leu Asn Arg Gln Ile	165	170
55	Cys Gly Ser Ser Ile Val Leu Leu Lys Asn Glu Asn Ser Ile Leu Pro	180	185
	Leu Pro Lys Ser Val Lys Lys Val Ala Leu Val Gly Ser His Val Arg	195	200
	Leu Pro Ala Ile Ser Gly Gly Gly Ser Ala Ser Leu Val Pro Tyr Tyr	210	215
	Ala Ile Ser Leu Tyr Asp Ala Val Ser Glu Val Leu Ala Gly Ala Thr	225	230
		235	240

Ile Thr His Glu Val Gly Ala Tyr Ala His Gln Met Leu Pro Val Ile  
 245 250 255  
 5

Asp Ala Met Ile Ser Asn Ala Val Ile His Phe Tyr Asn Asp Pro Ile  
 260 265 270

10

Asp Val Lys Asp Arg Lys Leu Leu Gly Ser Glu Asn Val Ser Ser Thr  
 275 280 285

15

Ser Phe Gln Leu Met Asp Tyr Asn Asn Ile Pro Thr Leu Asn Lys Ala  
 290 295 300

20

Met Phe Trp Gly Thr Leu Val Gly Glu Phe Ile Pro Thr Ala Thr Gly  
 305 310 315 320

25

Ile Trp Glu Phe Gly Leu Ser Val Phe Gly Thr Ala Asp Leu Tyr Ile  
 325 330 335

30

Asp Asn Glu Leu Val Ile Glu Asn Thr Thr His Gln Thr Arg Gly Thr  
 340 345 350

35

Ala Phe Phe Gly Lys Gly Thr Thr Glu Lys Val Ala Thr Arg Arg Met  
 355 360 365

40

Val Ala Gly Ser Thr Tyr Lys Leu Arg Leu Glu Phe Gly Ser Ala Asn  
 370 375 380

45

Thr Thr Lys Met Glu Thr Thr Gly Val Val Asn Phe Gly Gly Gly Ala  
 385 390 395 400

Val His Leu Gly Ala Cys Leu Lys Val Asp Pro Gln Glu Met Ile Ala  
 405 410 415

50

Arg Ala Val Lys Ala Ala Ala Asp Ala Asp Tyr Thr Ile Ile Cys Thr  
 420 425 430

55

Gly Leu Ser Gly Glu Trp Glu Ser Glu Gly Phe Asp Arg Pro His Met  
 435 440 445

Asp Leu Pro Pro Gly Val Asp Thr Met Ile Ser Gln Val Leu Asp Ala  
 450 455 460

Ala Pro Asn Ala Val Val Val Asn Gln Ser Gly Thr Pro Val Thr Met  
 465 470 475 480

5 Ser Trp Ala His Lys Ala Lys Ala Ile Val Gln Ala Trp Tyr Gly Gly  
 485 490 495

10 Asn Glu Thr Gly His Gly Ile Ser Asp Val Leu Phe Gly Asn Val Asn  
 500 505 510

15 Pro Ser Gly Lys Leu Ser Leu Ser Trp Pro Val Asp Val Lys His Asn  
 515 520 525

20 Pro Ala Tyr Leu Asn Tyr Ala Ser Val Gly Gly Arg Val Leu Tyr Gly  
 530 535 540

Glu Asp Val Tyr Val Gly Tyr Lys Phe Tyr Asp Lys Thr Glu Arg Glu  
 545 550 555 560

25 Val Leu Phe Pro Phe Gly His Gly Leu Ser Tyr Ala Thr Phe Lys Leu  
 565 570 575

30 Pro Asp Ser Thr Val Arg Thr Val Pro Glu Thr Phe His Pro Asp Gln  
 580 585 590

35 Pro Thr Val Ala Ile Val Lys Ile Lys Asn Thr Ser Ser Val Pro Gly  
 595 600 605

40 Ala Gln Val Leu Gln Leu Tyr Ile Ser Ala Pro Asn Ser Pro Thr His  
 610 615 620

Arg Pro Val Lys Glu Leu His Gly Phe Glu Lys Val Tyr Leu Glu Ala  
 625 630 635 640

45 Gly Glu Glu Lys Glu Val Gln Ile Pro Ile Asp Gln Tyr Ala Thr Ser  
 645 650 655

50 Phe Trp Asp Glu Ile Glu Ser Met Trp Lys Ser Glu Arg Gly Ile Tyr  
 660 665 670

55 Asp Val Leu Val Gly Phe Ser Ser Gln Glu Ile Ser Gly Lys Gly Lys  
 675 680 685

Leu Ile Val Pro Glu Thr Arg Phe Trp Met Gly Leu  
 690 695 700

5 <210> 93  
 <211> 1496  
 <212> DNA  
 <213> Artificial Sequence

10 <220>  
 <223> Maize optimized CBHI

<400> 93  
 tgcagtcgcg ctgcaccctc cagtccgaga cccaccgcgc gctcacctgg cagaagtgct 60  
 15 cctccggcgg cacctgcacc cagcagaccg gctccgtggg gatcgacgcc aactggcgct 120  
 ggaccacgc caccaactcc tccaccaact gctacgacgg caacacctgg tcctccaccc 180  
 20 tctgcccgga caacgagacc tgcgccaaaga actgtgcctt cgacggcgcc gcctacgcct 240  
 ccacctacgg cgtgaccacc tccggcaact cctctccat cggttcgtg acccagtcgg 300  
 cccagaagaa cgtgggcgc cgcctctacc tcatggcctc cgacaccacc taccaggagt 360  
 25 tcacctcctt cgccaacgag ttctccttcg acgtggacgt gtcccagctc ccgtgcggcc 420  
 tcaacggcgc cctctacttc gtgtccatgg acgccgacgg cggcgtgtcc aagtaccgga 480  
 30 ccaacaccgc cggcgccaag tacggcaccg gctactgcga ctcccagtgc ccgcgcgacc 540  
 tcaagtccat caacggccag gccaacgtgg agggctggga gccgtcctcc aacaacgcc 600  
 acaccggcat cggcggccac ggctcctgct gctccgagat ggacatctgg gaggccaact 660  
 35 ccatctcga ggccctcacc ccgcaccctg gcaccaccgt gggccaggag atctgcgagg 720  
 gcgacggctg cggcggcacc tactccgaca accgctacgg cggcacctgc gaccgggacg 780  
 40 gctgcgactg gaaccctgac cgcctcggca acacctcctt ctacggcccg ggctcctcct 840  
 tcacctcga caccaccaag aagctcaccg tggtagacca gtctgagacc tccggcgcca 900  
 tcaaccgcta ctacgtgcag aacggcgtga ccttcagca gccgaacgcc gagctcggct 960  
 45 cctactccgg caacgagctc aacgacgact actgcaccgc cgaggaggcc gagttcggcg 1020  
 gctcctcctt ctccgacaag ggcggcctca ccagttcaa gaaggccacc tccggcgcca 1080  
 50 tgggtgctcg gatgtccctc tgggacgact actacgcca catgctctgg ctcgactcca 1140  
 cctacccgac caacgagacc tctccaccc cgggcgcctg gcgcggctcc tgctccacct 1200  
 cctccggcgt gccggcccag gtggagtccc agtccccgaa cgccaagggtg accttctcca 1260  
 55 acatcaagtt cggcccgatc ggctccaccg gcaaccctgc cggcgccaac ccgcccggcg 1320

gcaaccgcggc gggcaccacc accaccgcggc gcccgggccac caccaccggc tcctccccgg 1380  
 gcccgaccca gtccactac ggccagtgcg gcggcatcgg ctactccggc ccgaccgtgt 1440  
 5 ggcctccgg caccacctgc caggtgctca acccgacta ctcccagtgc ctctag 1496

<210> 94  
 <211> 1365  
 10 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Maize optimized CBHII

15 <400> 94  
 atggtgcgcg tcgaggagcg ccaggcctgc tcctccgtgt ggggccagtg cggcggccag 60  
 aactggtcgg gcccgacctg ctgcgcctcc ggctccacct gcgtgtactc caacgactac 120  
 20 tactcccagt gcctcccggg cgccgcctcc tcctcctcct ccaccgcgcg cgcctccacc 180  
 acctcccgcg tgtcccgac cacctcccg cctcctccg ccaccgcgc gccgggctcc 240  
 25 accaccaacc gcgtgccgccc ggtgggctcc ggcaccgcca cctactccgg caaccggttc 300  
 gtgggcgtga ccccgtaggc caacgcctac tacgcctccg aggtgtctc cctcgccatc 360  
 ccgtccctca ccggcgccat ggccaccgcc gccgcggccg tggccaaggt gccgtccttc 420  
 30 atgtggctcg acaccctcga caagaccccg ctcatggagc agaccctcgc cgacatccgc 480  
 accgccaaca agaacggcgg caactacgcc ggccagttcg tgggtgtacga cctcccgac 540  
 35 cgcgactgcg ccgcccctgc ctccaacggc gactactcca tcgccgacgg cggcgtggcc 600  
 aagtacaaga actacatcga caccatccgc cagatcgtgg tggagtactc cgacatccgc 660  
 accctcctcg tgatcgagcc ggactccctc gccaacctcg tgaccaacct cggcaccgcc 720  
 40 aagtgcgcca acgcccagtc cgctacctc gactgcatca actacgccgt gaccagctc 780  
 aacctccga acgtggccat gtacctcgac gccggccacg ccggctggct cggctggccg 840  
 45 gccaaaccagg acccgggcgc ccagctcttc gccaacgtgt acaagaacgc tcctccccg 900  
 cgcgccctcc gcggcctcgc caccaacgtg gccaaactaca acggctggaa catcacctcc 960  
 ccgccgtcct acaccaggg caacgccgtg tacaacgaga agctctacat ccacgccatc 1020  
 50 ggcccgtcct tcgccaaaca cggctggctc aacgccttct tcatcaccga ccagggccgc 1080  
 tccggcaagc agccgaccgg ccagcagcag tggggcgact ggtgcaacgt gatcggcacc 1140  
 55 ggcttcggca tccgcccgtc cgccaacacc ggcgactccc tcctcgactc ctctgtgtgg 1200  
 gtgaagccgg gcggcgagtg cgacggcacc tcgactcct ccgccccgcg ctctgactcc 1260

cactgcgccc tcccggacgc cctccagccg gccccgcagg ccggcgccg gttccaggcc 1320  
 tacttcgtgc agctcctcac caacgccaac ccgtccttcc tctag 1365  
 5  
 <210> 95  
 <211> 1317  
 <212> DNA  
 10 <213> Artificial Sequence  
 <220>  
 <223> Maize optimized EGLI  
 15 <400> 95  
 atgcagcagc cgggcacctc caccocggag gtgcacccga agctcaccac ctacaagtgc 60  
 accaagtccg gcggtgcgt ggcccaggac acctccgtgg tgctcgactg gaactaccgc 120  
 20 tggatgcacg acgccaacta caactcctgc accgtgaacg gcggcggtgaa caccaccctc 180  
 tgcccggacg aggccacctg cggcaagaac tgcttcatcg agggcggtgga ctacgccgcc 240  
 tccggcggtga ccacctccgg ctctcctc accatgaacc agtacatgcc gtctcctcc 300  
 25 ggcggtact cctccgtgtc ccgcgcctc tacctcctcg actccgacgg cgagtacgtg 360  
 atgctcaagc tcaacggcca ggagctctcc ttcgacgtgg acctctccgc cctcccgtgc 420  
 30 ggcgagaacg gctccctcta cctctcccag atggacgaga acggcgggcg caaccagtac 480  
 aacaccgccg gcgccaacta cggtccggc tactgcgacg ccagtgccc ggtgcagacc 540  
 tggcgcaacg gcacctcaa cacctccac cagggttct gctgcaacga gatggacatc 600  
 35 ctcgagggca actcccgcg caacgcctc accccgact cctgcaccgc caccgcctgc 660  
 gactccgccg gctgcggctt caaccgtac ggctccggtt acaagtcta ctacggccccg 720  
 40 ggcgacaccg tggacacctc caagacctc accatcatca ccagttcaa caccgacaac 780  
 ggctccccgt ccggcaacct cgtgtccatc acccgcaagt accagcagaa cggcgtggac 840  
 atcccgtccg ccagccggg cggcgacacc atctcctct gcccgtccgc ctccgcctac 900  
 45 ggcggcctcg ccacatggg caaggccctc tctccggca tgggtgctgt gttctccatc 960  
 tggaaacgaca actcccagta catgaactgg ctcgactccg gcaacgcgg cccgtgctcc 1020  
 50 tccaccgagg gcaaccgctc caacaccctc gccaacaacc cgaacacca cgtggtgttc 1080  
 tccaacatcc gctggggcga catcggtctc accaccaact ccaccgccc gccgcccgcg 1140  
 ccggcctctt ccaccacctt ctccaccacc cgcgctctt ccaccacctc ctctcccccg 1200  
 55 tctgcaccc agaccactg ggccagtg ggcggcatcg gctactccg ctgcaagacc 1260

tgcacctcgg gcaccacctg ccagtactcc aacgactact actcccagtg cctctag 1317

<210> 96  
 5 <211> 1401  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 10 <223> Maize optimized BGLII

<400> 96  
 atgctcccga aggacttcca gtggggcttc gccaccgccc cctaccagat cgagggcgcc 60  
 15 gtggaccagg acggcccgcg ccgctccatc tgggacacct tctgcgcca gccgggcaag 120  
 atcgccgacg gtcctccgg cgtgaccgcc tgcgactcct acaaccgcac cgccgaggac 180  
 atcgccctcc tcaagtccct cggcgccaag tctaccgct tctccatctc ctgggtcccg 240  
 20 atcatcccg agggcgccg cggcgacgcc gtgaaccagg ccggcatcga ccactacgtg 300  
 aagtctgtgg acgacctct cgacgcgggc atcacccgt tcatcaccct cttccactgg 360  
 25 gacctcccg agggcctcca ccagcgctac ggcggcctcc tcaaccgcac cgagttcccg 420  
 ctcgacttcg agaactacgc ccgctgatg ttccgcgccc tcccgaagg ggcgaactgg 480  
 atcaccttca acgagccgct ctgctccgcc atcccgggct acggctccgg caccttcgcc 540  
 30 ccgggcccgc agtccacctc cgagccgtgg accgtgggcc acaacatcct cgtggccac 600  
 ggccgcgcg tgaaggccta ccgcgacgac ttcaagccgg cctccggcga cggccagatc 660  
 35 ggcacgtgc tcaacggcga cttcacctac ccgtgggacg ccgccgaccc ggccgacaag 720  
 gaggccgcg agcgccgcct cgagttcttc accgcctggt tcgccgaccc gatctacctc 780  
 ggcgactacc cggcctccat gcgcaagcag ctccggcgacc gcctcccgac cttcaccccc 840  
 40 gaggagcgcg cctcgtgca cggctccaac gacttctacg gcatgaacca ctacacctcc 900  
 aactacatcc gccaccgctc ctccccggcc tccgccgacg acaccgtggg caacgtggac 960  
 45 gtgctcttca ccaacaagca gggcaactgc atcgccccgg agaccagtc cccgtggctc 1020  
 cgcccgtgcg ccgcgggctt ccgcgacttc ctctgtgtga tctccaagcg ctacggctac 1080  
 ccgccgatct acgtgaccga gaacggcacc tccatcaagg gcgagtcga cctcccgaag 1140  
 50 gagaagatcc tcgaggacga cttccgcgtg aagtactaca acgagtacat ccgcgccatg 1200  
 gtgaccgcg tggagctcga cggcgtgaac gtgaagggt acttcgcctg gtccctcatg 1260  
 55 gacaacttcg agtgggcca cggctacgtg acccgcttcg gcgtgacct cgtggactac 1320  
 gagaacggcc agaagcgctt cccgaagaag tccgccaagt ccctcaagcc gctcttcgac 1380



gagctcatcg ccgccgccta g

1401

5 <210> 97  
 <211> 2103  
 <212> DNA  
 <213> Artificial Sequence

10 <220>  
 <223> Maize optimized CEL3D

<400> 97  
 15 atgacacctcg gctgcgagtc caccggcgtg atctccgccg tgaagcactt cgtggccaac 60  
 gaccaggagc acgagcgccg cgccgtggac tgcctcatca cccagcgcg cctccgcgag 120  
 gtgtacctec gcccgttcca gatcgtggcc cgcgacgcc gcccgggcgc cctcatgacc 180  
 20 tectacaaca aggtgaacgg caagcacgtg gccgactccg ccgagttcct ccagggcatac 240  
 ctccgcaccg agtggaaactg ggaccgcgtc atcgtgtccg actggtacgg cacctacacc 300  
 accatcgacg ccatcaaggc cggcctcgac ctcgagatgc cgggcgtgtc ccgctaccgc 360  
 25 ggcaagtaca tcgagtccgc cctccaggcc cgcctcctca agcagtcac catcgacgag 420  
 cgcgcccgcc gcggtgtccg ctccgccag aaggcctccc acctcaagg gtccgaggtg 480  
 30 gagcagggcc gcgacttccc ggaggaccgc gtgtcaacc gccagatctg cggtcctcc 540  
 atcgtgtctc tcaagaacga gaactccatc ctcccgctcc cgaagtccgt gaagaagggtg 600  
 gccctcgtgg gctccacgt gcgcctcccg gccatctccg gcggcggtc cgcctccctc 660  
 35 gtgccgtact acgccatctc cctctacgac gccgtgtccg aggtgctcgc cggcgccacc 720  
 atcacccacg aggtggggcg ctacgccac cagatgtctc cggtgatcga cgccatgac 780  
 40 tccaacgccg tgatccactt ctacaacgac ccgatcgacg tgaaggaccg caagctctc 840  
 ggctccgaga acgtgtctc cacctcctc cagctcatgg actacaaca catcccgacc 900  
 ctcaacaagg ccatgttctg gggcaccctc gtgggcgagt tcatcccgac cgccaccggc 960  
 45 atctgggagt tcggcctctc cgtgttcggc accgccgacc tctacatcga caacgagctc 1020  
 gtgatcgaga acaccacca ccagaccgc ggcaccgct tcttcggcaa gggcaccacc 1080  
 50 gagaagggtg ccaccgccg catggtggcc ggctccacct acaagctccg cctcgagttc 1140  
 ggctccgcca acaccacaa gatggagacc accggcgtgg tgaacttcgg cggcggcgcc 1200  
 gtgcacctcg gcgcctgcct caaggtggac ccgcaggaga tgatcgccc cgccgtgaag 1260  
 55 gccgcgcgcg acgccgacta caccatcatc tgcaccggcc tctccggcga gtgggagtcc 1320

	gagggcttcg accgcccga catggacctc ccgcccggcg tggacaccat gatctcccag	1380
	gtgctcgacg ccgccccgaa cgccgtgggtg gtgaaccagt ccggcacccc ggtgaccatg	1440
5	tcctggggccc acaaggccaa ggccatcgtg caggcctggg acggcggcaa cgagaccggc	1500
	cacggcatct ccgacgtgct ctteggcaac gtgaaccctg ccggcaagct ctccctctcc	1560
	tggccggtgg acgtgaagca caaccgggc tacctcaact acgcctcctg gggcggccgc	1620
10	gtgctctacg gcgaggacgt gtacgtgggc tacaagttct acgacaagac cgagcgcgag	1680
	gtgctcttcc cgttcggcca cggcctctcc tacgccacct tcaagctccc ggactccacc	1740
15	gtgcgcaccg tgccggagac ctccacccg gaccagccga ccgtggccat cgtgaagatc	1800
	aagaacacct cctccgtgcc gggcgcccag gtgctccagc tctacatctc cgcgccgaac	1860
	tccccgaccc accgcccgtt gaaggagctc cagggttcg agaagggtga cctcgaggcc	1920
20	ggcgaggaga aggaggtgca gatcccgatc gaccagtacg ccacctcctt ctgggacgag	1980
	atcgagtcca tgtggaagtc cgagcgcggc atctacgacg tgctcgtggg cttctctctc	2040
25	caggagatct ccggcaaggg caagctcatc gtgccggaga cccgcttctg gatgggcctc	2100
	tag	2103
30	<210> 98	
	<211> 420	
	<212> DNA	
	<213> Zea mays	
35	<220>	
	<223> Q protein promoter	
	<400> 98	
40	gggctggtaa attacttggg agcaatggta tgcaaatcct ttgcatgtac gcaaaactag	60
	ctagtgtgca caagttgtat atcgattcgt cgcgtttcaa caactcatgc aacattacaa	120
	acaagtaaca caatattaca aagttagttt catacaaagc aagaaaagga caataatact	180
45	tgacatgtaa agtgaagctt attatacttc ctaatccaac acaaaacaaa aaaaagttgc	240
	acaaagggtc aaaaatccac atcaaccatt aacctatacg taaagtgagt gatgagtcac	300
	attatccaac aaatgtttat caatgtggta tcatacaagc attgacatcc cataaatgca	360
50	agaaattgtg ccaacaaagc tataagtaac cctcatatgt atttgactc atgcatcaca	420
	<210> 99	
55	<211> 1188	
	<212> DNA	
	<213> artificial sequence	

&lt;220&gt;

&lt;223&gt; synthetic ferulic acid esterase

5 <400> 99  
 atggccgcct ccctcccgac catgccgcgc tccggctacg accaggtgcg caacggcgctg 60  
 ccgcgcggcc aggtgggtgaa catctectac ttctccaccg ccaccaactc caccgccccg 120  
 10 gcccgcgtgt acctcccgcc gggctactcc aaggacaaga agtactccgt gctctacctc 180  
 ctccacggca tcggcggtc cgagaacgac tggttcgagg gcggcgggcg cgccaacgtg 240  
 atcgccgaca acctcatgc cgagggcaag atcaagccgc tcatcatcgt gaccccgaa 300  
 15 accaacgccc ccggccccgg catcgccgac ggctacgaga acttcaccaa ggacctctc 360  
 aactccctca tcccgtaac cgagtccaac tactccgtgt acaccgaccg cgagcaccgc 420  
 20 gccatcgccg gcctctctat gggcgggcgc cagtccttca acatcgccct caccaacctc 480  
 gacaagtctg cctacatcg cccgatctcc gccgccccga acacctacc gaacgagcgc 540  
 ctcttcccg acggcgga ggcgccccgc gagaagctca agtcctctt catcgccctg 600  
 25 ggcaccaacg actccctcat cggcttcggc cagcgcgtgc acgagtactg cgtggccaac 660  
 aacatcaacc acgtgtactg gctcatccag ggcggcggcc acgacttcaa cgtgtggaag 720  
 30 ccgggcctct ggaacttct ccagatggcc gacgaggccg gcctcaccg cgacggcaac 780  
 accccggtgc cgaccccgtc ccgaagccg gccaacaccc gcatcgaggc cgaggactac 840  
 gacggcatca actcctctc catcgagatc atcggcgtgc cgccggaggg cggcccgggc 900  
 35 atcggtaca tcacctccg cgactacctc gtgtacaagt ccatcgactt cggcaacggc 960  
 gccacctct tcaaggccaa ggtggccaac gccaacacct ccaacatcga gcttcgcctc 1020  
 40 aacggccccg acggcaccct catcggcacc ctctccgtga agtccaccg cgactggaac 1080  
 acctacgagg agcagacctg ctccatctcc aagggtgacc gcatcaacga cctctacctc 1140  
 gtgttcaagg gcccggtgaa catcgactgg ttcaccttcg gcgtgtag 1188  
 45

&lt;210&gt; 100

&lt;211&gt; 395

&lt;212&gt; PRT

50 &lt;213&gt; artificial sequence

&lt;220&gt;

&lt;223&gt; synthetic ferulic acid esterase

55 &lt;400&gt; 100

Met Ala Ala Ser Leu Pro Thr Met Pro Pro Ser Gly Tyr Asp Gln Val

1	5	10	15
5	Arg Asn Gly Val Pro Arg Gly Gln Val Val Asn Ile Ser Tyr Phe Ser	20	25
10	Thr Ala Thr Asn Ser Thr Arg Pro Ala Arg Val Tyr Leu Pro Pro Gly	35	40
15	Tyr Ser Lys Asp Lys Lys Tyr Ser Val Leu Tyr Leu Leu His Gly Ile	50	55
20	Gly Gly Ser Glu Asn Asp Trp Phe Glu Gly Gly Gly Arg Ala Asn Val	65	70
25	Ile Ala Asp Asn Leu Ile Ala Glu Gly Lys Ile Lys Pro Leu Ile Ile	85	90
30	Val Thr Pro Asn Thr Asn Ala Ala Gly Pro Gly Ile Ala Asp Gly Tyr	100	105
35	Glu Asn Phe Thr Lys Asp Leu Leu Asn Ser Leu Ile Pro Tyr Ile Glu	115	120
40	Ser Asn Tyr Ser Val Tyr Thr Asp Arg Glu His Arg Ala Ile Ala Gly	130	135
45	Leu Ser Met Gly Gly Gly Gln Ser Phe Asn Ile Gly Leu Thr Asn Leu	145	150
50	Asp Lys Phe Ala Tyr Ile Gly Pro Ile Ser Ala Ala Pro Asn Thr Tyr	165	170
55	Pro Asn Glu Arg Leu Phe Pro Asp Gly Gly Lys Ala Ala Arg Glu Lys	180	185
	Leu Lys Leu Leu Phe Ile Ala Cys Gly Thr Asn Asp Ser Leu Ile Gly	195	200
	Phe Gly Gln Arg Val His Glu Tyr Cys Val Ala Asn Asn Ile Asn His	210	215
	Val Tyr Trp Leu Ile Gln Gly Gly Gly His Asp Phe Asn Val Trp Lys	225	230
		235	240

5 Pro Gly Leu Trp Asn Phe Leu Gln Met Ala Asp Glu Ala Gly Leu Thr  
 245 250 255  
 Arg Asp Gly Asn Thr Pro Val Pro Thr Pro Ser Pro Lys Pro Ala Asn  
 260 265 270  
 10 Thr Arg Ile Glu Ala Glu Asp Tyr Asp Gly Ile Asn Ser Ser Ser Ile  
 275 280 285  
 15 Glu Ile Ile Gly Val Pro Pro Glu Gly Gly Arg Gly Ile Gly Tyr Ile  
 290 295 300  
 20 Thr Ser Gly Asp Tyr Leu Val Tyr Lys Ser Ile Asp Phe Gly Asn Gly  
 305 310 315 320  
 Ala Thr Ser Phe Lys Ala Lys Val Ala Asn Ala Asn Thr Ser Asn Ile  
 325 330 335  
 25 Glu Leu Arg Leu Asn Gly Pro Asn Gly Thr Leu Ile Gly Thr Leu Ser  
 340 345 350  
 30 Val Lys Ser Thr Gly Asp Trp Asn Thr Tyr Glu Glu Gln Thr Cys Ser  
 355 360 365  
 35 Ile Ser Lys Val Thr Gly Ile Asn Asp Leu Tyr Leu Val Phe Lys Gly  
 370 375 380  
 40 Pro Val Asn Ile Asp Trp Phe Thr Phe Gly Val  
 385 390 395  
 <210> 101  
 <211> 1188  
 45 <212> DNA  
 <213> artificial sequence  
 <220>  
 <223> plasmid 13036  
 50 <400> 101  
 atggccgcct ccctcccgac catgccgccg tccggctacg accaggtgcg caacggcggtg 60  
 ccgcgcggcc aggtggtgaa catctctac ttctccaccg ccaccaactc caccgccccg 120  
 55 gccgcggtgt acctcccgcc gggctactcc aaggacaaga agtactccgt gctctacctc 180

ctccacggca tcggcggtc cgagaacgac tggttcgagg gcggcgccg cgccaacgtg 240  
 atcgccgaca acctcatcgc cgagggcaag atcaagccgc tcatcatcgt gaccccgaa 300  
 5 accaacgccg ccggcccggg catcgccgac ggctacgaga acttcaccaa ggacctctc 360  
 aactcctca tcccgtaacat cgagtccaac tactcgtgt acaccgaccg cgagcaccgc 420  
 gccatcgccg gcctctctat gggcgccggc cagtccttca acatcggcct caccaacctc 480  
 10 gacaagttcg cctacatcgg ccgatctcc gccgccccga acacctacc gaacgagcgc 540  
 ctcttcccg acggcggaac ggccgcccgc gagaagctca agtcctctt catcgctgc 600  
 15 ggcaccaacg actcctcat cggcttcggc cagcgcgtgc acgagtactg cgtggccaac 660  
 aacatcaacc acgtgtactg gctcatccag ggcgccggcc acgacttcaa cgtgtggaag 720  
 ccgggcctct ggaacttct ccagatggcc gacgaggccg gcctcaccg cgacggcaac 780  
 20 acccgggtgc cgacccgc cccgaagccg gccaacacc gcacgaggc cgaggactac 840  
 gacggcatca actcctctc catcgagatc atcggcgtgc cgccggaggg cggccgcggc 900  
 25 atcggtaca tcacctcgg cgactacctc gtgtacaagt ccatcgactt cggcaacggc 960  
 gccacctct tcaaggccaa ggtggccaac gccaacacct ccaacatcga gcttcgcctc 1020  
 aacggccga acggcaccct catcggcacc ctctcgtga agtcaccgg cgactggaac 1080  
 30 acctacgagg agcagacctg ctccatctcc aaggtagccg gcatcaacga cctctacctc 1140  
 gtgttcaagg gcccggtgaa catcgactgg ttcacctcg gcgtgtag 1188

35  
 <210> 102  
 <211> 395  
 <212> PRT  
 <213> artificial sequence

40  
 <220>  
 <223> plasmid 13036

45  
 <400> 102

Met Ala Ala Ser Leu Pro Thr Met Pro Pro Ser Gly Tyr Asp Gln Val  
 1 5 10 15

50 Arg Asn Gly Val Pro Arg Gly Gln Val Val Asn Ile Ser Tyr Phe Ser  
 20 25 30

Thr Ala Thr Asn Ser Thr Arg Pro Ala Arg Val Tyr Leu Pro Pro Gly  
 35 40 45

Tyr Ser Lys Asp Lys Lys Tyr Ser Val Leu Tyr Leu Leu His Gly Ile  
 50 55 60

5 Gly Gly Ser Glu Asn Asp Trp Phe Glu Gly Gly Gly Arg Ala Asn Val  
 65 70 75 80

10 Ile Ala Asp Asn Leu Ile Ala Glu Gly Lys Ile Lys Pro Leu Ile Ile  
 85 90 95

Val Thr Pro Asn Thr Asn Ala Ala Gly Pro Gly Ile Ala Asp Gly Tyr  
 100 105 110

15 Glu Asn Phe Thr Lys Asp Leu Leu Asn Ser Leu Ile Pro Tyr Ile Glu  
 115 120 125

20 Ser Asn Tyr Ser Val Tyr Thr Asp Arg Glu His Arg Ala Ile Ala Gly  
 130 135 140

25 Leu Ser Met Gly Gly Gly Gln Ser Phe Asn Ile Gly Leu Thr Asn Leu  
 145 150 155 160

30 Asp Lys Phe Ala Tyr Ile Gly Pro Ile Ser Ala Ala Pro Asn Thr Tyr  
 165 170 175

35 Pro Asn Glu Arg Leu Phe Pro Asp Gly Gly Lys Ala Ala Arg Glu Lys  
 180 185 190

Leu Lys Leu Leu Phe Ile Ala Cys Gly Thr Asn Asp Ser Leu Ile Gly  
 195 200 205

40 Phe Gly Gln Arg Val His Glu Tyr Cys Val Ala Asn Asn Ile Asn His  
 210 215 220

45 Val Tyr Trp Leu Ile Gln Gly Gly Gly His Asp Phe Asn Val Trp Lys  
 225 230 235 240

50 Pro Gly Leu Trp Asn Phe Leu Gln Met Ala Asp Glu Ala Gly Leu Thr  
 245 250 255

Arg Asp Gly Asn Thr Pro Val Pro Thr Pro Ser Pro Lys Pro Ala Asn  
 260 265 270

55 Thr Arg Ile Glu Ala Glu Asp Tyr Asp Gly Ile Asn Ser Ser Ser Ile

	275	280	285	
5	Glu Ile Ile Gly Val Pro	Pro Glu Gly Gly Arg	Gly Ile Gly Tyr Ile	
	290	295	300	
10	Thr Ser Gly Asp Tyr Leu Val Tyr Lys Ser Ile Asp Phe Gly Asn Gly			
	305	310	315	320
15	Ala Thr Ser Phe Lys Ala Lys Val Ala Asn Ala Asn Thr Ser Asn Ile			
		325	330	335
20	Glu Leu Arg Leu Asn Gly Pro Asn Gly Thr Leu Ile Gly Thr Leu Ser			
		340	345	350
25	Val Lys Ser Thr Gly Asp Trp Asn Thr Tyr Glu Glu Gln Thr Cys Ser			
		355	360	365
30	Ile Ser Lys Val Thr Gly Ile Asn Asp Leu Tyr Leu Val Phe Lys Gly			
		370	375	380
35	Pro Val Asn Ile Asp Trp Phe Thr Phe Gly Val			
		385	390	395
40	<210> 103			
	<211> 1245			
	<212> DNA			
	<213> artificial sequence			
45	<220>			
	<223> plasmid 13038			
50	<400> 103			
	atgaggggtgt tgctcggttgc cctcgctctc ctggtctctcg ctgcgagcgc cacctccatg			60
	gccgcctccc tcccgaccat gccgcccgtcc ggctacgacc aggtgcgcaa cggcgtgccg			120
55	cgcgccagg tggtgaacat ctctacttc tccaccgcca ccaactccac ccgcccggcc			180
	cgcgtgtacc tcccgcggg ctactccaag gacaagaagt actccgtgct ctacctctc			240
	cacggcatcg gcggctccga gaacgactgg ttcgagggcg gcggccgcgc caacgtgac			300
	gccgacaacc tcatcgccga gggcaagatc aagccgctca tcatcgtgac cccgaacacc			360
	aacgccgcg gcccgggcat cgccgacggc tacgagaact tcaccaagga cctcctcaac			420
	ttcctcatcc cgtacatcga gtccaactac tccgtgtaca ccgaccgca gcaccgcgc			480
	atcgccggcc tctctatggg cggcggccag tccttcaaca tcggcctcac caacctcgac			540



aagttcgctt acatcgcccc gatctccgcc gccccgaaca cctaccgaa cgagcgcttc 600  
 tccccggacg gcggcaaggc cccccgcgag aagctcaagc tctcttcat cgcctgcggc 660  
 5 accaacgact ccctcatcgg ctccggccag cgcgtgcacg agtactgctt ggccaacaac 720  
 atcaaccacg tgtactggct catccagggc ggcgccacg acttcaacgt gtggaagccg 780  
 10 ggctcttggg acttctctca gatggccgac gagggcggcc tcaccgcga cggcaaacacc 840  
 ccggtgccga ccccgcccc gaagccggcc aacaccgca tcgaggccga ggactacgac 900  
 ggcatacaact cctctccat cgagatcatc ggcgcgccg cggaggcgcg ccgcggcatc 960  
 15 ggctacatca cctccggcga ctacctctg tacaagtcca tcgacttcgg caacggcgcc 1020  
 acctccttca aggccaaggt ggccaacgcc aacacctcca acatcgagct tcgctcaac 1080  
 20 ggcccgaacg gcacctcat cggcacctc tccgtgaagt ccaccggcga ctggaacacc 1140  
 tacgaggagc agacctgct catctccaag gtgaccggca tcaacgacct ctacctctg 1200  
 ttcaagggcc cggatgaacat cgactgggtc accttcggcg tctag 1245  
 25

<210> 104

<211> 414

<212> PRT

30 <213> artificial sequence

<220>

<223> plasmid 13038 aa

35 <400> 104

Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser  
 1 5 10 15

40 Ala Thr Ser Met Ala Ala Ser Leu Pro Thr Met Pro Pro Ser Gly Tyr  
 20 25 30

45 Asp Gln Val Arg Asn Gly Val Pro Arg Gly Gln Val Val Asn Ile Ser  
 35 40 45

50 Tyr Phe Ser Thr Ala Thr Asn Ser Thr Arg Pro Ala Arg Val Tyr Leu  
 50 55 60

Pro Pro Gly Tyr Ser Lys Asp Lys Lys Tyr Ser Val Leu Tyr Leu Leu  
 65 70 75 80

55

His Gly Ile Gly Gly Ser Glu Asn Asp Trp Phe Glu Gly Gly Gly Arg

	85	90	95
5	Ala Asn Val Ile Ala Asp Asn Leu Ile Ala Glu Gly Lys Ile Lys Pro 100 105 110		
10	Leu Ile Ile Val Thr Pro Asn Thr Asn Ala Ala Gly Pro Gly Ile Ala 115 120 125		
15	Asp Gly Tyr Glu Asn Phe Thr Lys Asp Leu Leu Asn Ser Leu Ile Pro 130 135 140		
20	Tyr Ile Glu Ser Asn Tyr Ser Val Tyr Thr Asp Arg Glu His Arg Ala 145 150 155 160		
25	Ile Ala Gly Leu Ser Met Gly Gly Gly Gln Ser Phe Asn Ile Gly Leu 165 170 175		
30	Thr Asn Leu Asp Lys Phe Ala Tyr Ile Gly Pro Ile Ser Ala Ala Pro 180 185 190		
35	Asn Thr Tyr Pro Asn Glu Arg Leu Phe Pro Asp Gly Gly Lys Ala Ala 195 200 205		
40	Arg Glu Lys Leu Lys Leu Leu Phe Ile Ala Cys Gly Thr Asn Asp Ser 210 215 220		
45	Leu Ile Gly Phe Gly Gln Arg Val His Glu Tyr Cys Val Ala Asn Asn 225 230 235 240		
50	Ile Asn His Val Tyr Trp Leu Ile Gln Gly Gly Gly His Asp Phe Asn 245 250 255		
55	Val Trp Lys Pro Gly Leu Trp Asn Phe Leu Gln Met Ala Asp Glu Ala 260 265 270		
60	Gly Leu Thr Arg Asp Gly Asn Thr Pro Val Pro Thr Pro Ser Pro Lys 275 280 285		
65	Pro Ala Asn Thr Arg Ile Glu Ala Glu Asp Tyr Asp Gly Ile Asn Ser 290 295 300		
70	Ser Ser Ile Glu Ile Ile Gly Val Pro Pro Glu Gly Gly Arg Gly Ile 305 310 315 320		

Gly Tyr Ile Thr Ser Gly Asp Tyr Leu Val Tyr Lys Ser Ile Asp Phe  
 325 330 335  
 5  
 Gly Asn Gly Ala Thr Ser Phe Lys Ala Lys Val Ala Asn Ala Asn Thr  
 340 345 350  
 10  
 Ser Asn Ile Glu Leu Arg Leu Asn Gly Pro Asn Gly Thr Leu Ile Gly  
 355 360 365  
 15 Thr Leu Ser Val Lys Ser Thr Gly Asp Trp Asn Thr Tyr Glu Glu Gln  
 370 375 380  
 20 Thr Cys Ser Ile Ser Lys Val Thr Gly Ile Asn Asp Leu Tyr Leu Val  
 385 390 395 400  
 Phe Lys Gly Pro Val Asn Ile Asp Trp Phe Thr Phe Gly Val  
 405 410  
 25  
 <210> 105  
 <211> 1425  
 <212> DNA  
 30 <213> artificial sequence  
 <220>  
 <223> plasmid 13039  
 35 <400> 105  
 atgctggcgg ctctggccac gtcgcagctc gtcgcaacgc gcgccggcct gggcggtccc 60  
 gacgcgtcca cggtccgccg cggcgccgcg cagggcctga gggggggccc ggcgtcgccg 120  
 40 gcggcggaca cgctcagcat gcggaccagc gcgcgcgcgg cggccaggca ccagcaccag 180  
 caggcgcgcc gcggggccag gttcccgtcg ctctcgtgtg gcgccagcgc cggcgccatg 240  
 gccgcctccc tcccgaccat gccgccgtcc ggctacgacc aggtgcgcaa cggcgtgccg 300  
 45 cgcggccagg tggatgaacat ctctacttc tccaccgcca ccaactccac ccgcccgccc 360  
 cgcgtgtacc tcccgccggg ctactccaag gacaagaagt actccgtgct ctacctctc 420  
 50 cacggcatcg gcggctccga gaacgactgg ttcgagggcg gcggccgcgc caacgtgatc 480  
 gccgacaacc tcatcgccga gggcaagatc aagccgctca tcatcgtgac cccgaacacc 540  
 aacgcgcgcg gcccgggcat cgcgcagcgc tacgagaact tcaccaagga cctcctcaac 600  
 55 tccctcatcc cgtacatcga gtccaactac tccgtgtaca ccgaccgcga gcaccgcgcc 660

atcgccggcc tctctatggg cggcggccag tccttcaaca tcggcctcac caacctcgac 720  
 aagttcgctt acatcgggcc gatctccgcc gccccgaaca cctacccgaa cgagcgcctc 780  
 5 tccccggacg gcggaaggc cggccgcgag aagctcaagc tcctcttcat cgcctgcggc 840  
 accaacgact cctcatcgg ctctggccag cgcgtgcacg agtactgctt ggccaacaac 900  
 atcaaccacg tgtactggct catccagggc ggcgccacg acttcaacgt gtggaagccg 960  
 10 ggctcttga acttctcca gatggccgac gaggccggcc tcaccgcga cggcaacacc 1020  
 ccggtgccga ccccgctccc gaagccggcc aacaccgca tcgaggccga ggactacgac 1080  
 15 ggcatcaact cctcctccat cgagatcatc ggctgcccgc cggagggcgg ccgcggcatc 1140  
 ggctacatca cctccggcga ctacctcgtg tacaagtcca tcgacttcgg caacggcgcc 1200  
 acctccttca aggccaaggt ggccaacgcc aacacctcca acatcgagct tcgcctcaac 1260  
 20 gggccgaacg gcacctcat cggcacctc tccgtgaagt ccaccggcga ctggaacacc 1320  
 tacgaggagc agacctgctc catctccaag gtgaccggca tcaacgacct ctacctcgtg 1380  
 25 ttcaagggcc cgggtgaacat cgactgggtc accttcggcg tgtag 1425

<210> 106  
 <211> 474  
 30 <212> PRT  
 <213> artificial sequence

<220>  
 <223> plasmid 13039 aa

35 <400> 106

Met Leu Ala Ala Leu Ala Thr Ser Gln Leu Val Ala Thr Arg Ala Gly  
 1 5 10 15  
 40 Leu Gly Val Pro Asp Ala Ser Thr Phe Arg Arg Gly Ala Ala Gln Gly  
 20 25 30  
 45 Leu Arg Gly Ala Arg Ala Ser Ala Ala Ala Asp Thr Leu Ser Met Arg  
 35 40 45  
 50 Thr Ser Ala Arg Ala Ala Pro Arg His Gln His Gln Gln Ala Arg Arg  
 50 55 60  
 55 Gly Ala Arg Phe Pro Ser Leu Val Val Cys Ala Ser Ala Gly Ala Met  
 65 70 75 80

Ala Ala Ser Leu Pro Thr Met Pro Pro Ser Gly Tyr Asp Gln Val Arg  
85 90 95

5 Asn Gly Val Pro Arg Gly Gln Val Val Asn Ile Ser Tyr Phe Ser Thr  
100 105 110

10 Ala Thr Asn Ser Thr Arg Pro Ala Arg Val Tyr Leu Pro Pro Gly Tyr  
115 120 125

15 Ser Lys Asp Lys Lys Tyr Ser Val Leu Tyr Leu Leu His Gly Ile Gly  
130 135 140

Gly Ser Glu Asn Asp Trp Phe Glu Gly Gly Gly Arg Ala Asn Val Ile  
145 150 155 160

20 Ala Asp Asn Leu Ile Ala Glu Gly Lys Ile Lys Pro Leu Ile Ile Val  
165 170 175

25 Thr Pro Asn Thr Asn Ala Ala Gly Pro Gly Ile Ala Asp Gly Tyr Glu  
180 185 190

30 Asn Phe Thr Lys Asp Leu Leu Asn Ser Leu Ile Pro Tyr Ile Glu Ser  
195 200 205

35 Asn Tyr Ser Val Tyr Thr Asp Arg Glu His Arg Ala Ile Ala Gly Leu  
210 215 220

Ser Met Gly Gly Gly Gln Ser Phe Asn Ile Gly Leu Thr Asn Leu Asp  
225 230 235 240

40 Lys Phe Ala Tyr Ile Gly Pro Ile Ser Ala Ala Pro Asn Thr Tyr Pro  
245 250 255

45 Asn Glu Arg Leu Phe Pro Asp Gly Gly Lys Ala Ala Arg Glu Lys Leu  
260 265 270

50 Lys Leu Leu Phe Ile Ala Cys Gly Thr Asn Asp Ser Leu Ile Gly Phe  
275 280 285

Gly Gln Arg Val His Glu Tyr Cys Val Ala Asn Asn Ile Asn His Val  
290 295 300

55 Tyr Trp Leu Ile Gln Gly Gly Gly His Asp Phe Asn Val Trp Lys Pro

	305		310		315		320	
5	Gly	Leu	Trp	Asn	Phe	Leu	Arg	
				325		330	335	
10	Asp	Gly	Asn	Thr	Pro	Val	Thr	
			340		345		350	
15	Arg	Ile	Glu	Ala	Glu	Asp	Tyr	
		355			360		365	
20	Ile	Ile	Gly	Val	Pro	Pro	Glu	
		370			375		380	
25	Ser	Gly	Asp	Tyr	Leu	Val	Tyr	
	385				390		395	
30	Thr	Ser	Phe	Lys	Ala	Lys	Val	
				405		410		
35	Leu	Arg	Leu	Asn	Gly	Pro	Asn	
			420		425		430	
40	Lys	Ser	Thr	Gly	Asp	Trp	Asn	
			435		440		445	
45	Ser	Lys	Val	Thr	Gly	Ile	Asn	
	450				455		460	
50	Val	Asn	Ile	Asp	Trp	Phe	Thr	
	465				470			
55	<210> 107							
	<211> 1263							
	<212> DNA							
	<213> artificial sequence							
	<220>							
	<223> plasmid 13347							
	<400> 107							
	atgagggtgt tgctcggtgc cctcgctctc ctggctctcg ctgcgagcgc cacctccatg							60
	gccgcctccc tcccgaccat gccgcgctcc ggctacgacc aggtgcgcaa cggcgtgccg							120
	cgcggccagg tggatgaacat ctctacttc tccaccgcca ccaactccac ccgcccggcc							180

cgcggtgtacc tcccgcggg ctactccaag gacaagaagt actccgtgct ctacctctc 240  
 cacggcatcg gcggctccga gaacgactgg ttcgagggcg gcggccgcgc caacgtgatc 300  
 5 gccgacaacc tcacgcccga gggcaagatc aagccgctca tcacgtgac cccgaacacc 360  
 aacgcgcgcg gcccgggcat cgccgacggc tacgagaact tcaccaagga cctcctcaac 420  
 10 tccctcatcc cgtacatcga gtccaactac tccgtgtaca ccgaccgcca gcaccgcgc 480  
 atcgccggcc tctctatggg cgggggccag tccttcaaca tcggcctcac caacctcgac 540  
 aagttcgctt acatcgcccc gatctccgcc gccccgaaca cctaccgaa cgagcgctc 600  
 15 ttcccggacg gcggcaaggc cgcccgcgag aagctcaagc tcctcttcat cgctgcggc 660  
 accaacgact cctcatcgg cttcgccag cgcggtgacg agtactgctt ggccaacaac 720  
 20 atcaaccacg tgtactggct catccagggc ggcgccacg acttcaacgt gtggaagccg 780  
 ggcctctgga acttctcca gatggccgac gaggccggcc tcaccgcca cggaacacc 840  
 ccggtgccga cccgtcccc gaagccggcc aacaccgca tcgaggcca ggactacgac 900  
 25 ggcatacaact cctctccat cgagatcatc ggcggtgccg cgaggggcg ccgcggcatc 960  
 ggctacatca cctccggcca ctacctgtg tacaagtcca tcgacttcgg caacggcgcc 1020  
 30 acctcttca aggccaaggt ggccaacgcc aacacctcca acatcgagct tcgctcaac 1080  
 ggcccgaacg gcacctcat cggcaccctc tccgtgaagt ccaccggcca ctggaacacc 1140  
 tacgaggagc agacctgtc catctccaag gtgaccggca tcaacgacct ctacctgtg 1200  
 35 ttcaagggcc cggtgaacat cgactgggtc accttcggcg tgtccgagaa ggacgaactc 1260  
 tag 1263

40  
 <210> 108  
 <211> 420  
 <212> PRT  
 <213> artificial sequence

45  
 <220>  
 <223> plasmid 13347

50  
 <400> 108

Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser  
 1 5 10 15

55 Ala Thr Ser Met Ala Ala Ser Leu Pro Thr Met Pro Pro Ser Gly Tyr  
 20 25 30

Asp Gln Val Arg Asn Gly Val Pro Arg Gly Gln Val Val Asn Ile Ser  
           35                          40                          45  
 5  
 Tyr Phe Ser Thr Ala Thr Asn Ser Thr Arg Pro Ala Arg Val Tyr Leu  
           50                          55                          60  
 10 Pro Pro Gly Tyr Ser Lys Asp Lys Lys Tyr Ser Val Leu Tyr Leu Leu  
       65                          70                          75                          80  
 15 His Gly Ile Gly Gly Ser Glu Asn Asp Trp Phe Glu Gly Gly Gly Arg  
                           85                          90                          95  
 20 Ala Asn Val Ile Ala Asp Asn Leu Ile Ala Glu Gly Lys Ile Lys Pro  
                           100                          105                          110  
 Leu Ile Ile Val Thr Pro Asn Thr Asn Ala Ala Gly Pro Gly Ile Ala  
           115                          120                          125  
 25 Asp Gly Tyr Glu Asn Phe Thr Lys Asp Leu Leu Asn Ser Leu Ile Pro  
           130                          135                          140  
 30 Tyr Ile Glu Ser Asn Tyr Ser Val Tyr Thr Asp Arg Glu His Arg Ala  
       145                          150                          155                          160  
 35 Ile Ala Gly Leu Ser Met Gly Gly Gly Gln Ser Phe Asn Ile Gly Leu  
                           165                          170                          175  
 40 Thr Asn Leu Asp Lys Phe Ala Tyr Ile Gly Pro Ile Ser Ala Ala Pro  
                           180                          185                          190  
 Asn Thr Tyr Pro Asn Glu Arg Leu Phe Pro Asp Gly Gly Lys Ala Ala  
           195                          200                          205  
 45 Arg Glu Lys Leu Lys Leu Leu Phe Ile Ala Cys Gly Thr Asn Asp Ser  
       210                          215                          220  
 50 Leu Ile Gly Phe Gly Gln Arg Val His Glu Tyr Cys Val Ala Asn Asn  
       225                          230                          235                          240  
 55 Ile Asn His Val Tyr Trp Leu Ile Gln Gly Gly Gly His Asp Phe Asn  
                           245                          250                          255



Asp Gln Val Arg Asn Gly Val Pro Arg Gly Gln Val Val Asn Ile Ser  
 35 40 45  
 5  
 Tyr Phe Ser Thr Ala Thr Asn Ser Thr Arg Pro Ala Arg Val Tyr Leu  
 50 55 60  
 10 Pro Pro Gly Tyr Ser Lys Asp Lys Lys Tyr Ser Val Leu Tyr Leu Leu  
 65 70 75 80  
 15 His Gly Ile Gly Gly Ser Glu Asn Asp Trp Phe Glu Gly Gly Gly Arg  
 85 90 95  
 20 Ala Asn Val Ile Ala Asp Asn Leu Ile Ala Glu Gly Lys Ile Lys Pro  
 100 105 110  
 Leu Ile Ile Val Thr Pro Asn Thr Asn Ala Ala Gly Pro Gly Ile Ala  
 115 120 125  
 25 Asp Gly Tyr Glu Asn Phe Thr Lys Asp Leu Leu Asn Ser Leu Ile Pro  
 130 135 140  
 30 Tyr Ile Glu Ser Asn Tyr Ser Val Tyr Thr Asp Arg Glu His Arg Ala  
 145 150 155 160  
 35 Ile Ala Gly Leu Ser Met Gly Gly Gly Gln Ser Phe Asn Ile Gly Leu  
 165 170 175  
 40 Thr Asn Leu Asp Lys Phe Ala Tyr Ile Gly Pro Ile Ser Ala Ala Pro  
 180 185 190  
 Asn Thr Tyr Pro Asn Glu Arg Leu Phe Pro Asp Gly Gly Lys Ala Ala  
 195 200 205  
 45 Arg Glu Lys Leu Lys Leu Leu Phe Ile Ala Cys Gly Thr Asn Asp Ser  
 210 215 220  
 50 Leu Ile Gly Phe Gly Gln Arg Val His Glu Tyr Cys Val Ala Asn Asn  
 225 230 235 240  
 55 Ile Asn His Val Tyr Trp Leu Ile Gln Gly Gly Gly His Asp Phe Asn  
 245 250 255

Val Trp Lys Pro Gly Leu Trp Asn Phe Leu Gln Met Ala Asp Glu Ala  
                     260                    265                    270

5 Gly Leu Thr Arg Asp Gly Asn Thr Pro Val Pro Thr Pro Ser Pro Lys  
                     275                    280                    285

10 Pro Ala Asn Thr Arg Ile Glu Ala Glu Asp Tyr Asp Gly Ile Asn Ser  
                     290                    295                    300

15 Ser Ser Ile Glu Ile Ile Gly Val Pro Pro Glu Gly Gly Arg Gly Ile  
                     305                    310                    315                    320

Gly Tyr Ile Thr Ser Gly Asp Tyr Leu Val Tyr Lys Ser Ile Asp Phe  
                     325                    330                    335

20 Gly Asn Gly Ala Thr Ser Phe Lys Ala Lys Val Ala Asn Ala Asn Thr  
                     340                    345                    350

25 Ser Asn Ile Glu Leu Arg Leu Asn Gly Pro Asn Gly Thr Leu Ile Gly  
                     355                    360                    365

30 Thr Leu Ser Val Lys Ser Thr Gly Asp Trp Asn Thr Tyr Glu Glu Gln  
                     370                    375                    380

35 Thr Cys Ser Ile Ser Lys Val Thr Gly Ile Asn Asp Leu Tyr Leu Val  
                     385                    390                    395                    400

40 Phe Lys Gly Pro Val Asn Ile Asp Trp Phe Thr Phe Gly Val Ser Glu  
                     405                    410                    415

Lys Asp Glu Leu  
                     420

45 <210> 109  
       <211> 1296  
       <212> DNA  
       <213> artificial sequence

50 <220>  
       <223> plasmid 11267

55 <400> 109  
       atgaggggtgt tgctcgttgc cctcgctctc ctggctctcg ctgcgagcgc caccagcgct 60  
       gcgcagtcgc agccggagct gaagctggag tccgtggtga tcgtgtcccg ccacggcgctg 120

cgcgccccga ccaaggccac ccagctcatg caggacgtga ccccgagcgc ctggccgacc 180  
 tggccggtga agctcggcga gctgaccccg cgcggcggcg agctgatcgc ctacctcggc 240  
 5 cactactggc gccagcgctt cgtggccgac ggctcctcc cgaagtgcgg ctgcccgcag 300  
 tccggccagg tggccatcat cgccgacgtg gacgagcgca cccgcaagac cggcgaggcc 360  
 ttgcgcgcgc gcctcgcccc ggactgcgcc atcaccgtgc acaccaggc cgacacctcc 420  
 10 tccccggacc cgctcttcaa cccgctcaag accggcgtgt gccagctcga caacgccaac 480  
 gtgaccgacg ccatacctgga gcgcgcgcgc ggctccatcg ccgacttcac cggccactac 540  
 15 cagaccgcct tccgcgagct ggagcgcggt ctcaacttcc cgcagtccaa cctctgcctc 600  
 aagcgcgaga agcaggacga gtctgtctcc ctacccagg ccctcccgtc cgagctgaag 660  
 gtgtccgcgc actgcgtgtc cctcaccggc gccgtgtccc tcgcctccat gctcaccgaa 720  
 20 atcttctctc tccagcaggc ccagggcacg ccggagccgg gctggggccg catcaccgac 780  
 tcccaccagt ggaacacctt cctctccctc cacaacgccc agttcgacct cctccagcgc 840  
 25 accccggagg tggcccgcgc ccgcgcaccc ccgctcctcg acctcatcaa gaccgcctc 900  
 accccgcacc cgccgcagaa gcaggcctac ggcgtgaccc tcccgaacct cgtgctcttc 960  
 atcgccggcc acgacaccaa cctcgccaac ctcggcggcg ccctggagct gaactggacc 1020  
 30 ctcccgggccc agccggacaa caccgcgcgc ggcggcgagc tgggtgttcga gcgctggcgc 1080  
 cgcctctccg acaactccca gtggattcag gtgtccctcg tgttccagac cctccagcag 1140  
 35 atgcgcgaca agaccccgct ctccctcaac accccgcgcg gcgaggtgaa gctcaccctc 1200  
 gccggctgcg aggagcgcaa cgcccagggc atgtgctccc tcgcccggctt caccagatc 1260  
 gtgaacgagg cccgcacccc ggctgtctcc ctctaa 1296  
 40

<210> 110

<211> 431

<212> PRT

45 <213> artificial sequence

<220>

<223> plasmid 11267 aa sequence

50 <400> 110

Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser  
 1 5 10 15

55

Ala Thr Ser Ala Ala Gln Ser Glu Pro Glu Leu Lys Leu Glu Ser Val  
 20 25 30

Val Ile Val Ser Arg His Gly Val Arg Ala Pro Thr Lys Ala Thr Gln  
 35 40 45  
 5  
 Leu Met Gln Asp Val Thr Pro Asp Ala Trp Pro Thr Trp Pro Val Lys  
 50 55 60  
 10  
 Leu Gly Glu Leu Thr Pro Arg Gly Gly Glu Leu Ile Ala Tyr Leu Gly  
 65 70 75 80  
 15 His Tyr Trp Arg Gln Arg Leu Val Ala Asp Gly Leu Leu Pro Lys Cys  
 85 90 95  
 20 Gly Cys Pro Gln Ser Gly Gln Val Ala Ile Ile Ala Asp Val Asp Glu  
 100 105 110  
 Arg Thr Arg Lys Thr Gly Glu Ala Phe Ala Ala Gly Leu Ala Pro Asp  
 115 120 125  
 25 Cys Ala Ile Thr Val His Thr Gln Ala Asp Thr Ser Ser Pro Asp Pro  
 130 135 140  
 30 Leu Phe Asn Pro Leu Lys Thr Gly Val Cys Gln Leu Asp Asn Ala Asn  
 145 150 155 160  
 35 Val Thr Asp Ala Ile Leu Glu Arg Ala Gly Gly Ser Ile Ala Asp Phe  
 165 170 175  
 40 Thr Gly His Tyr Gln Thr Ala Phe Arg Glu Leu Glu Arg Val Leu Asn  
 180 185 190  
 Phe Pro Gln Ser Asn Leu Cys Leu Lys Arg Glu Lys Gln Asp Glu Ser  
 195 200 205  
 45 Cys Ser Leu Thr Gln Ala Leu Pro Ser Glu Leu Lys Val Ser Ala Asp  
 210 215 220  
 50 Cys Val Ser Leu Thr Gly Ala Val Ser Leu Ala Ser Met Leu Thr Glu  
 225 230 235 240  
 55 Ile Phe Leu Leu Gln Gln Ala Gln Gly Met Pro Glu Pro Gly Trp Gly  
 245 250 255

Arg Ile Thr Asp Ser His Gln Trp Asn Thr Leu Leu Ser Leu His Asn  
 260 265 270  
 5  
 Ala Gln Phe Asp Leu Leu Gln Arg Thr Pro Glu Val Ala Arg Ser Arg  
 275 280 285  
 10 Ala Thr Pro Leu Leu Asp Leu Ile Lys Thr Ala Leu Thr Pro His Pro  
 290 295 300  
 15 Pro Gln Lys Gln Ala Tyr Gly Val Thr Leu Pro Thr Ser Val Leu Phe  
 305 310 315 320  
 20 Ile Ala Gly His Asp Thr Asn Leu Ala Asn Leu Gly Gly Ala Leu Glu  
 325 330 335  
 Leu Asn Trp Thr Leu Pro Gly Gln Pro Asp Asn Thr Pro Pro Gly Gly  
 340 345 350  
 25 Glu Leu Val Phe Glu Arg Trp Arg Arg Leu Ser Asp Asn Ser Gln Trp  
 355 360 365  
 30 Ile Gln Val Ser Leu Val Phe Gln Thr Leu Gln Gln Met Arg Asp Lys  
 370 375 380  
 35 Thr Pro Leu Ser Leu Asn Thr Pro Pro Gly Glu Val Lys Leu Thr Leu  
 385 390 395 400  
 40 Ala Gly Cys Glu Glu Arg Asn Ala Gln Gly Met Cys Ser Leu Ala Gly  
 405 410 415  
 Phe Thr Gln Ile Val Asn Glu Ala Arg Ile Pro Ala Cys Ser Leu  
 420 425 430  
 45  
 <210> 111  
 <211> 1314  
 <212> DNA  
 <213> artificial sequence  
 50  
 <220>  
 <223> plasmid 11268  
 <400> 111  
 55 atgagggtgt tgctcggtgc cctcgctctc ctggctctcg ctgcgagcgc caccagcgct 60  
 gcgcagtcgc agccggagct gaagctggag tccgtggtga tcgtgtcccg ccacggcgctg 120

cgcgccccga ccaaggccac ccagctcatg caggacgtga ccccgagcgc ctggccgacc 180  
 5 tggecggtga agctcggcga gctgaccccg cgcgggcgcg agctgacgc ctacctcggc 240  
 cactactggc gccagcgccct cgtggccgac ggctcctcc cgaagtgcgg ctgcccgcag 300  
 tccggccagg tggccatcat cgccgacgtg gacgagcgca cccgcaagac cggcgaggcc 360  
 10 ttcgccgccc gcctcgcccc ggactgcgcc atcaccgtgc acaccaggc cgacacctcc 420  
 tccccggacc cgctcttcaa cccgctcaag accggcggtgt gccagctcga caacgccaac 480  
 gtgaccgacg ccattcctgga gcgcgcgggc ggctccatcg ccgacttcac cggccactac 540  
 15 cagaccgcct tccgcgagct ggagcgcggtg ctcaacttcc cgcagtccaa cctctgcctc 600  
 aagcgcgaga agcaggacga gtctgtctcc ctacccagg ccctcccgtc cgagctgaag 660  
 20 gtgtccgccc actgcgtgtc cctcaccggc gccgtgtccc tcgctccat gctcaccgaa 720  
 atcttctctc tccagcaggc ccagggcatg ccggagcgcg gctggggccg catcaccgac 780  
 tcccaccagt ggaacacct cctctcctc cacaacgcc agttcgacct cctccagcgc 840  
 25 accccggagg tggcccgtc ccgcgccacc ccgtctctc acctcatcaa gaccgcctc 900  
 accccgcacc cgccgcagaa gcaggcctac ggcgtagacc tcccgcctc cgtgctcttc 960  
 30 atcgccggcc acgacaccaa cctcgccaac ctcggcggcg ccctggagct gaactggacc 1020  
 ctcccgggccc agccggacaa caccgcgcg ggcggcgagc tgggtgttcga gcgctggcgc 1080  
 cgctctctcg acaactccca gtggattcag gtgtccctcg tgttcagac cctccagcag 1140  
 35 atgcgcgaca agaccgcgt ctccctcaac acccgcccg gcgaggtgaa gctcaccctc 1200  
 gccggctgcg aggagcgcaa cgcccaggc atgtgtctcc tcgccggtt caccagatc 1260  
 40 gtgaacgagg cccgcatccc ggcctgtctc ctctccgaga aggacgagct gtaa 1314

<210> 112  
 <211> 437  
 45 <212> PRT  
 <213> artificial sequence  
 <220>  
 <223> plasmid 11268 amino acid sequence  
 50 <400> 112

Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser  
 1 5 10 15  
 55  
 Ala Thr Ser Ala Ala Gln Ser Glu Pro Glu Leu Lys Leu Glu Ser Val

	20	25	30
5	Val Ile Val Ser Arg His Gly Val Arg Ala Pro Thr Lys Ala Thr Gln 35 40 45		
10	Leu Met Gln Asp Val Thr Pro Asp Ala Trp Pro Thr Trp Pro Val Lys 50 55 60		
15	Leu Gly Glu Leu Thr Pro Arg Gly Gly Glu Leu Ile Ala Tyr Leu Gly 65 70 75 80		
20	His Tyr Trp Arg Gln Arg Leu Val Ala Asp Gly Leu Leu Pro Lys Cys 85 90 95		
25	Gly Cys Pro Gln Ser Gly Gln Val Ala Ile Ile Ala Asp Val Asp Glu 100 105 110		
30	Arg Thr Arg Lys Thr Gly Glu Ala Phe Ala Ala Gly Leu Ala Pro Asp 115 120 125		
35	Cys Ala Ile Thr Val His Thr Gln Ala Asp Thr Ser Ser Pro Asp Pro 130 135 140		
40	Leu Phe Asn Pro Leu Lys Thr Gly Val Cys Gln Leu Asp Asn Ala Asn 145 150 155 160		
45	Val Thr Asp Ala Ile Leu Glu Arg Ala Gly Gly Ser Ile Ala Asp Phe 165 170 175		
50	Thr Gly His Tyr Gln Thr Ala Phe Arg Glu Leu Glu Arg Val Leu Asn 180 185 190		
55	Phe Pro Gln Ser Asn Leu Cys Leu Lys Arg Glu Lys Gln Asp Glu Ser 195 200 205		
60	Cys Ser Leu Thr Gln Ala Leu Pro Ser Glu Leu Lys Val Ser Ala Asp 210 215 220		
65	Cys Val Ser Leu Thr Gly Ala Val Ser Leu Ala Ser Met Leu Thr Glu 225 230 235 240		
70	Ile Phe Leu Leu Gln Gln Ala Gln Gly Met Pro Glu Pro Gly Trp Gly 245 250 255		

Arg Ile Thr Asp Ser His Gln Trp Asn Thr Leu Leu Ser Leu His Asn  
 260 265 270  
 5  
 Ala Gln Phe Asp Leu Leu Gln Arg Thr Pro Glu Val Ala Arg Ser Arg  
 275 280 285  
 10  
 Ala Thr Pro Leu Leu Asp Leu Ile Lys Thr Ala Leu Thr Pro His Pro  
 290 295 300  
 15  
 Pro Gln Lys Gln Ala Tyr Gly Val Thr Leu Pro Thr Ser Val Leu Phe  
 305 310 315 320  
 20  
 Ile Ala Gly His Asp Thr Asn Leu Ala Asn Leu Gly Gly Ala Leu Glu  
 325 330 335  
 25  
 Leu Asn Trp Thr Leu Pro Gly Gln Pro Asp Asn Thr Pro Pro Gly Gly  
 340 345 350  
 30  
 Glu Leu Val Phe Glu Arg Trp Arg Arg Leu Ser Asp Asn Ser Gln Trp  
 355 360 365  
 35  
 Thr Pro Leu Ser Leu Asn Thr Pro Pro Gly Glu Val Lys Leu Thr Leu  
 385 390 395 400  
 40  
 Ala Gly Cys Glu Glu Arg Asn Ala Gln Gly Met Cys Ser Leu Ala Gly  
 405 410 415  
 45  
 Phe Thr Gln Ile Val Asn Glu Ala Arg Ile Pro Ala Cys Ser Leu Ser  
 420 425 430  
 50  
 Glu Lys Asp Glu Leu  
 435